UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

DEPARTAMENTO DE FÍSICA ATÓMICA, MOLECULAR Y NUCLEAR

TESIS DOCTORAL

Search for heavy Higgs-like resonances in the H→ZZ→$1^{+}1^{-}$qq final state in *pp* collisions in the CMS experimented at the LHC

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Daniel Domínguez Vázquez

Directores

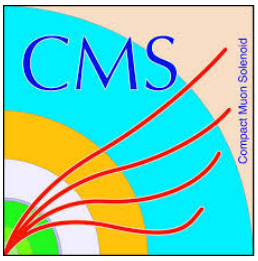
Mary-Cruz Fouz Iglesias
Óscar González López

Madrid, 2014

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas
(CIEMAT-Madrid)

•

Departamento de Física Atómica, Molecular y Nuclear
(Universidad Complutense de Madrid)



Search for heavy Higgs-like resonances in the $H \rightarrow ZZ \rightarrow l^+l^-q\bar{q}$ final state in *pp* collisions in the CMS Experiment at the LHC

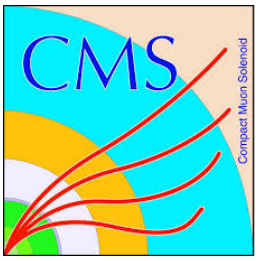
Thesis submited by
Daniel Domínguez Vázquez

Supervised by
Dra. Mary-Cruz Fouz Iglesias
Dr. Óscar González López

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas
(CIEMAT-Madrid)

•

Departamento de Física Atómica, Molecular y Nuclear
(Universidad Complutense de Madrid)



Búsqueda de resonancias masivas similares al bosón de Higgs, en el estado final $H \rightarrow ZZ \rightarrow l^+l^-q\bar{q}$ en colisiones *pp* en el experimento CMS en el LHC

Tesis presentada por
Daniel Domínguez Vázquez

Dirigida por
Dra. Mary-Cruz Fouz Iglesias
Dr. Óscar González López

Ós meus pais.
A Pepa e a Octavia.

Contents

List of Figures

List of Tables

Chapter 1

Summary

The Standard Model describes three of the four known interactions among fundamental particles, and it has been confirmed over the last decades in an innumerable amount of experiments. Only the mechanism to understand the masses of the fermions and weak bosons remained unsolved. The Brout-Englert-Higgs mechanism proposed a spontaneous symmetry breaking of an additional field. This mechanism implies the existence of a scalar boson, usually known as the "*Higgs boson*", of unknown mass, which had been unsuccessfully searched in LEP and Tevatron accelerators.

The Large Hadron Collider at CERN is the largest accelerator ever built, and one of its more important goals was the discovery of the Higgs boson, in case it existed. During 2012, the CMS and ATLAS Collaborations announced the discovery of a scalar boson of $m_H \sim 125$ GeV, which later was confirmed to be a Higgs boson. However, this is only the beginning of the understanding of the scalar sector and the mass generation of the fundamental particles. The pending questions about the Standard Model allow many models that predict additional resonances, very similar to those expected from the Higgs boson, and at a higher mass. Consequently, the study of Higgs signatures at higher masses is very important for the CMS Collaboration in order to confirm or reject these models.

In order to achieve relevant results, an accurate and efficient detector is needed. The Compact Muon Solenoid has shown an impressive performance since the first collisions in 2009. The best possible detection and determination of muons, which are involved in many processes of interest, is vital. In consequence, the understanding of the behaviour of the different subdetectors involved is mandatory to proceed with the Physics analysis. The Drift Tube Chambers are responsible of triggering and measuring the trajectory of muons in the central part of CMS. They have been designed, built and maintained by the CIEMAT team. In the present thesis a study of the performance of the drift tube chambers is presented. Efficiency, resolution and noise contamination has been determined with the first data of collisions recorded, showing an excellent performance, in agreement with the design expectations.

In the study of other Higgs-like models, the $H \to ZZ \to l^+l^-q\bar{q}$ final state is a very promising final state, which combines a relatively high production rate with an acceptable background contamination. Furthermore, it allows to reconstruct the invariant mass m_{llqq} to look for resonances, and a powerful test statistics is made base on the shape of this distribution.

This thesis presents the analysis performed looking for heavy Higgs-like signatures in the $H \to ZZ \to l^+l^-q\bar{q}$ final state in the range 230-600 GeV, with the data recorded by CMS from 2010 to 2013. The data correspond to two different running periods: during 2010 and 2011 protons collided at an energy of $\sqrt{s}$ =7 TeV in the center of mass, while during 2012 and 2013 they collided at $\sqrt{s}$ =8 TeV. Due to the different conditions of the accelerator and the detector, two independent analyses have been performed, although they share most of the strategy. The latter, which is more sensitive to a Higgs signal, is where I contributed to a larger extend, and therefore is described in more detailed.

The analysis requires a deep understanding of the objects reconstructed by the CMS detector, in order to optimize the selection of events. Many processes, different from the one under study, leave the same signature in the detector, and in many cases they are produced several orders of magnitude more frequently than the Higgs, so a hard effort has been put to filter the sample and improve the signal expected fraction in the selected events.

Next, all the processes considered have been determined, and all the sources of systematic uncertainty have been evaluated.

Finally, results are presented under the interpretation of a Standard Model Higgs. It has been excluded the existence of such particle in the range 275-600 GeV at 95% CL. This result is interpreted as a benchmark to other models Beyond Standard Model. The fact that no indication of any process apart from the known background processes have been observed induces interesting conclusions in them. For the coming future, the next steps to improve this analysis include the extension of the range under study up to 1 TeV, a vector-boson-fusion production dedicated analysis and the study of other theoretical models.

Chapter 2

Introduction

"The Cosmos is all that is, or ever was, or ever will be", Carl Sagan stated at the beginning of his TV show. The apparent harmony and beauty of the Nature has motivated mankind over the History to wonder about the laws that rule it. Last century witnessed the major breakthrough in the understanding of our Universe, with the development of the Relativity, Quantum Mechanics and Quantum Field Theory. The Standard Model (SM) is the most accurate and general framework we have come up with for the moment. It describes three of the four known interactions among elementary particles, and it has been supported by a large number of observations.

One of the most elusive (from the experimental point of view) predictions of the SM was related with the explanation of the mass of the fundamental particles. The Brout-Englert-Higgs mechanism suggested a spontaneous symmetry breaking of an additional scalar field to solve the issue. The Large Hadron Collider (LHC) at CERN, and the Compact Muon Solenoid (CMS) experiment have been designed and constructed to solve many of the missing questions about the Standard Model, with special interest in the search of a scalar boson associated to that field, in order to reject or confirm this mass mechanism. The discovery of a scalar boson with mass $m_H \sim 125$ GeV, compatible to this moment with the SM prediction, by the CMS and ATLAS experiments in the LHC at CERN in 2012 has been one of the major discoveries of the last decades. However, the scalar sector is only starting to be understood, and the certainty that the Standard Model is not complete, originates many models which predict additional resonances with similar signatures. The experimental test of these models, like Supersymmetry, is also a task for the experiments of the LHC.

This thesis describes the search performed within the CMS Collaboration on any particle decaying into two Z bosons, one of them decays leptonically, and the other one hadronically: $H \to ZZ \to l^+l^-q\bar{q}$. The recorded data from 2010 to 2013 from proton-proton collisions produced by the LHC at an energy of $\sqrt{s}$ =7 TeV and $\sqrt{s}$ =8 TeV in the center of mass have been used. The integrated luminosity recorded in the first case is $\mathcal{L}$ =5.0 fb^{-1} and in the second case is of $\mathcal{L}$ =19.6 fb^{-1}. The $H \to ZZ \to l^+l^-q\bar{q}$ is a

very interesting final state to analyze, as it is sensitive to a range of mass hypothesis which had not been explored before the LHC. From the experimental point of view, it has the advantages that it has a production rate relatively large compared to another final states, and the possibility to fully reconstruct the invariant mass of the quadri-object. The results have been interpreted under the SM Higgs hypothesis, although the presence or not of any excess of data, or discrepancy with respect to the background prediction, may be an indication of New Physics beyond the Standard Model.

The work is presented as follows:

The chapter 3 summarizes the SM and the Brout-Englert-Higgs mechanism to explain the mass of the fundamental particles, including the expected production and decay modes of the Higgs boson. The results achieved to the moment of writing this thesis, on the discovery of a scalar boson with mass $m_H \sim 125$ GeV have also been included.

Chapter 4 briefly describes the LHC design and its performance during the first run of data production, up to February 2013. Chapter 5 in turn shows a comprehensive description of the CMS detector design and all the subsystems involved.

The CMS Collaboration owes its name to the effort put to identify and reconstruct muons with the best possible accuracy. Many of the processes of interest involve muons in its final state, including the $H \to ZZ \to l^+l^-q\bar{q}$, topic of this thesis. Hence, the system related to the muon detection is vital for CMS. Chapter 6 describes the design and performance of the Drift Tube Chambers (DT), responsible of triggering and measuring the muons crossing the central part of the detector. CIEMAT is responsible of the design, construction and maintenance of these chambers. The chapter focuses on the results obtained in the efficiency and resolution of the tubes, with the data recorded in the beginning of the LHC run. A good understanding of the chamber behaviour is important in the following reconstruction steps, and to improve the detector response. The modelling and influence of background in the drift tubes have also been analyzed.

The full reconstruction process and performance of the objects involved in the $H \to ZZ \to l^+l^-q\bar{q}$ final state are presented in chapter 7. These objects are muons, electrons, hadronic jets and missing energy. Additional quantities related to the characteristics of the objects, like isolation or b-tagging, are also detailed.

After all ingredients have been presented, the analysis performed on the $H \to ZZ \to l^+l^-q\bar{q}$ with the 8 TeV dataset recorded during 2012 and 2013 is detailed in chapter 8. This is the main chapter of this thesis. The features of this particular channel are explained. The overwhelmingly large production of a Z boson associated with hadronic jets from in QCD processes, known as Z+jets, is the largest background source to fight, five orders of magnitude larger than the expected signal presence. Other backgrounds involve the top-antitop production and the SM production of dibosons. An optimized selection of events is presented, in order to maximize the fraction of signal candidates in the sample. Then, the determination of the signal and backgrounds expectations is reported. A detailed description of the systematic uncertainty sources

and its determination follows, ended by the results obtained for this analysis. The prior analysis performed with the dataset recorded during 2011 from collisions at $\sqrt{s}$ =7 TeV follows a similar approach and many common features, although several improvements have been achieved, as the conditions of the data taking also varied. In consequence, only the differences between one and another are reported in chapter 9.

Finally, the statistical interpretation of the result, in terms of the SM Higgs hypothesis, is presented in chapter 10 and exclusion limits to the Higgs production are set. The SM interpretation must be considered only as a benchmark, and conclusions about other models may be extracted. And chapter 11 summarizes the work referred in the previous chapters, the conclusions obtained and the prospects to improve the results presented.

My contribution to this thesis includes, on one hand, the analysis on the performance of the DT chambers detailed in chapter 6, which was published in [2, 3]. On the other hand, I contributed to the two $H \to ZZ \to l^+l^-q\overline{q}$ analyses of collisions at 7 and 8 TeV, described in chapters 9 and 8, respectively. In the 7 TeV analysis I specially contributed to the statistical interpretation of the analysis and the limit calculation, although I also participated in the event selection and the validation of the simulation software. In the 8 TeV analysis my contribution was much more significant, participating in most of the features of the analysis: in the event selection, in the background and signal determination, and in the limit calculation. This work has been published in [4, 5, 6].

Natural units $\hbar$=c=1 will be used throughout the text.

I hope you enjoy.

Chapter 3

The Mass Mechanism in Particle Physics

The most accurate explanation of most of the observed phenomena in Nature, except gravitation, is given by the Standard Model of Particle Physics (SM) [7].

According to the Standard Model, matter is composed of a set of fundamental particles, described, along with its properties and interactions, in terms of quantum fields [8, 9]. It was developed through the 1960 and 1970 decades, and it is probably the most accurate theory developed by the human being, given the power of its predictions, which were confirmed with high precision in different experiments through the last decades in many experiments, and also led to the discovery of different particles (J/Ψ (1974)[10], W boson (1983)[11, 12] , Z boson (1983)[13, 14], Top quark (1995) [15, 16], τ neutrino (2000)[17] and finally the Higgs boson (2012)) [18, 19].

This chapter tries to summarize the concepts of the SM and the Higgs mechanism, including the last findings in the Higgs boson search, as well as the miscomprehended features in the present day and possible solutions.

3.1 The Standard Model of Particle Physics

The SM includes the existence of two kinds of particles, according to the value of their spin: matter is made of 1/2-spin particles called *fermions* which obey the Fermi-Dirac statistics, while their interactions are mediated by integer-spin particles called bosons, because they obey Bose-Einstein statistics instead. On top of that, there is another scalar called the Higgs-Englet-Brout boson that interacts with both fermions and bosons.

Fermions, in turn, are classified in *quarks*, which have *color* charge, and hence feel the strong force, and *leptons* which do not. Both quarks and leptons are found in three families. The first one is responsible of ordinary matter. The other two are replicas of the first one but with higher mass. Hence, they are unstable and will decay.

Each family of quarks is composed by a doublet: (u,d), (c,s), (t,b). Quarks have color and fractionary electric charge, so they interact through all types of forces (electromagnetic, weak and strong), whereas each family of leptons is formed by a negative charged particle and a neutrino. They do not have color charge, so they don't interact via strong force. Both are listed with their properties in Table 3.1 [1].

Fundamental forces are understood as interactions among particles through the exchange of mediators. These mediators are spin 1 bosons, called *photon* for the electromagnetic interaction, *gluon* for the strong and $W^{\pm}$,*Z bosons* for the weak interactions. They are listed in Table 3.2 [1]. Gravity is not considered in the Standard Model because it is negligible at quantum scales, and therefore we have no quantum theory for gravity yet. Nevertheless, a mediator is predicted as the carrier of gravity force, called *graviton*, with a mass m$< 7 \cdot 10^{-32} eV$ [1].

On top of that, for each particle exists a correspondent antiparticle, which has opposite quantum numbers but the same physical properties.

	Quarks				**Leptons**			
Generation	Name	Symbol	Mass	Charge	Name	Symbol	Mass	Charge
1^{st}	Up	u	2.3 MeV	2/3	Electron	e	0.51 MeV	-1
	Down	d	4.8 MeV	-1/3	Electron neutrino	ν_e	< 2 eV	0
2^{nd}	Charm	c	1.275 GeV	2/3	Muon	μ	105.7 MeV	-1
	Strange	s	95 MeV	-1/3	Muon neutrino	ν_μ	< 2 eV	0
3^{rd}	Top	t	173.07 GeV	2/3	Tau	τ	1.77 GeV	-1
	Bottom	b	4.66 GeV	-1/3	Tau neutrino	ν_τ	< 2 eV	0

Table 3.1: Table of quarks and leptons with their charge and mass, according to the Particle Data Group 2013 Review[1]

Interaction	Symbol	Mass	Charge
Electromagnetic	γ	0	0
Strong	g	0	0
Weak	W $^{\pm}$	80.38 GeV	± 1
Weak	Z	91.2 GeV	0

Table 3.2: Table of carrier bosons of the fundamental interactions.

The Standard Model is a gauge quantum field theory based on the symmetry group $\mathrm{SU(3)}_C \otimes \mathrm{SU(2)}_L \otimes \mathrm{U(1)}_Y$.

The fundamental interactions are derived from the principle of local gauge invariance with respect to this symmetry group. And according to Noether's theorem [20], any symmetry present in the system implies a conserved current. The dynamics of a system is derived from the principle of minimal action, by minimizing

$$S = \int \mathcal{L} dx^4 \tag{3.1}$$

where the corresponding Lagrangian is written in two terms, one including the strong sector, and the other the electroweak sector, which unifies electromagnetic and weak interactions.

$$\mathcal{L} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} \tag{3.2}$$

3.1.1 Quantum Electrodynamics (QED)

Quantum Electrodynamics is one of the most beautiful theories ever developed because it naturally arises simply from the principle of local gauge invariance.

Introducing any local internal phase transformation of the field α(x)

$$\psi(x) \rightarrow e^{i\alpha(x)}\psi(x) \tag{3.3}$$

in the relativistic Dirac Lagrangian describing the motion and mass of a free fermion

$$\mathcal{L} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi \tag{3.4}$$

makes necessary to replace the partial derivative in Eq. 3.4 by the covariant derivative, in order to conserve the invariance of the Lagrangian under this rotation:

$$D_{\mu} \equiv \partial_{\mu} - ieA_{\mu} \tag{3.5}$$

being e is the electric source and A_{μ} is a vector field transforming as

$$A_{\mu} \rightarrow A_{\mu} + \frac{1}{e}\partial_{\mu}\alpha \tag{3.6}$$

which leads to the Lagrangian

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi + e\bar{\psi}\gamma^{\mu}A_{\mu}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.7}$$

where the second term accounts for the interaction between the field A_{μ} and the electromagnetic source e of the field ψ, and the last term, where ${}_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ is the kinetic term of the electromagnetic field. Of course, from the Noether' s theorem, electric current is conserved and Maxwell' s equations are recoverable:

$$\partial_{\mu}F^{\mu\nu} = \mu_0 J^{\nu} \tag{3.8}$$

Finally, this same condition of gauge invariance is the one that forbids other terms like that for the mass of the photon.

3.1.2 Quantum Chronodynamics (QCD)

The strong force is responsible of the quarks holding together forming hadrons, like the proton or neutron. On top of that, the residual strong force is the responsible of the

different nucleons being held in the nuclei of the atoms. There are two particular features that characterize the strong force: the *confinement* and the *asymptotic freedom*. Confinement refers to the behaviour of strong interaction at low energies, which do not decrease with distance, which in turn explains why only mesons ($color + anticolor$) or baryons (red+blue+green) have been observed, and no free quarks or gluons. This is interpreted as the fact that only *colorless* objects are allowed by Nature, where color charge is an internal property of quarks, like electric charge. The difference with electric charge is that there exist three types of colors ($red, blue, green$) plus its anticolors ($\overline{red}, \overline{blue}, \overline{green}$). Asymptotic freedom explains how bonds at high energies (or equivalently, at small distances) become weaker, so particles become asymptotically free.

Mathematically, it is described analogously as in QED but noticing that three different color fields are considered instead of one electric field, so the gauge group will be SU(3) instead of U(1). So applying a local transformation over a quark field $q(x)$

$$q(x) \to e^{i\alpha_a(x)T_a} q(x) \tag{3.9}$$

$$D_\mu = \partial_\mu + ig_s T_a G^a \tag{3.10}$$

where T_a with a=1...8 are the generators of the symmetry group, g_s is the strong coupling and the fields G^a_μ introduced transform as

$$G^a_\mu \to G^a_\mu - \frac{1}{g}\partial_\mu \alpha_a - f_{abc}\alpha_b G^c_\mu \tag{3.11}$$

the Lagrangian obtained this way will be:

$$\mathcal{L} = \bar{q}(i\gamma^\mu \partial_\mu - m)q - g(\bar{q}\gamma^\mu T_a q)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{3.12}$$

where

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f_{abc} G^b_\mu G^c_\nu \tag{3.13}$$

is the gluon field tensor. The last term in Eq. 3.13, which does not exist in QED for being an abelian theory, includes the self-interaction between the gluon fields, through three and four gluon vertices, because gluons also carry color charge. And again, no mass term is allowed for gluons.

3.1.3 Electroweak unification

The attempts of describing the weak interaction, responsible of the "beta decay", with an analogous formalism as that shown before, led to the Glashow-Weinberg-Salam model, in which the unification with the electromagnetic interaction was needed at high energies, despite the obvious differences between both (weak is a short-distance interaction, and much *weaker* with respect to electromagnetism) [21, 22, 7]. Thus, a unified

$SU(2)_L \otimes U(1)_Y$ symmetry group was proposed to describe the weak and electromagnetic interactions together. At low energies, the symmetry is broken and the photon, and the massive $W^{\pm}$ and Z arise. The discovery of the latter at CERN was the final proof of this mechanism.

The experimental fact that only neutrinos with left chirality are observed in Nature [1], lead to the formulation of fermions in terms of a left part and right part, defined as the projections $\psi_{L,R} = 1/2(1 \mp \gamma^5))\psi$ of the fermion field and where only left part interacted weakly. Left fermions are doublets transforming under $SU(2)_L \otimes U(1)_Y$ and right fermions are singlets transforming only to $U(1)_Y$.

$$f_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, ... \qquad f_L \rightarrow f'_L = e^{\beta^a(x)T_a + i\alpha(x)Y} f_L \tag{3.14}$$

$$f_R = e_R, u_R, d_R, ... \qquad f_R \rightarrow f'_R = e^{i\alpha(x)Y} f_R \tag{3.15}$$

where $\beta^a(x)$ and $\alpha(x)$ are the local phases, T_a is the generator of $SU(2)_L$, and Y is the hypercharge, generator of $U(1)_Y$ and related with the charge and the weak isospin I_3:

$$Q = I_3 + \frac{Y}{2} \tag{3.16}$$

The corresponding covariant derivative results:

$$D_\mu = \partial_\mu + i\frac{g}{2}T_a W^i_\mu + i\frac{g'}{2}B_\mu Y \tag{3.17}$$

where g and g' are the couplings to W^i B respectively, so the electroweak Lagrangian remains:

$$\mathcal{L}_{EW} = \bar{\psi} i\gamma^\mu D_\mu \psi - \frac{1}{4}W^i_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{3.18}$$

where the first term includes the interaction between the fermions and the fields, which result, using Eq. 3.17 and Eq. 3.18

$$\mathcal{L}_{int} = -\overline{\psi_L}\gamma^\mu \left(\frac{g}{2}T_a W^i_\mu + \frac{g'}{2}B_\mu Y\right)\psi_L - \overline{\psi_R}\gamma^\mu \left(\frac{g'}{2}B_\mu Y\right)\psi_R \tag{3.19}$$

and

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon_{abc}W^b_\mu W^c_\nu \tag{3.20}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{3.21}$$

3.2 The Brout-Englert-Higgs Mass Mechanism

The previous formalism predicts the existence of 4 massless fields (W^i and B), which didn't agree with the expectation of massive boson fields for the weak force,

given its short range, and later in the 70's with the discovery of $W^{\pm}$ and Z bosons. Furthermore, fermions have to be massless, against observation.

Robert Brout and François Englert, at the same time as Peter Higgs, and based on the ideas of P.W. Anderson taken from the model of effective massive gauge bosons in superconductivity and plasmas, developed a mechanism to explain the masses of the mediator fields of electroweak interaction through a spontaneous symmetry breaking (SSB) via an additional field, nowadays known as the *Higgs field*, or *mass field* [23, 24, 25]. The formalism was further developed by T. Kibble, C.R. Hagen and G. Guralnik and finally included in the Electroweak model by Weinberg and Salam [26, 27, 28, 7]. The Brout-Englert-Higgs (BEH) mechanism is just the simplest choice for a spontaneous breaking of the symmetry preserving gauge invariance.

The idea of a SSB is that the Lagrangian is invariant under certain transformation, while the vacuum is not. So it is needed to introduce an additional field, which in the simplest case is a complex doublet field

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{3.22}$$

with a particular potential, like

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{3.23}$$

There are two possible minimum states depending on the sign of μ^2. If it is greater than zero, then the ground state is at $< \phi >= 0$ and the symmetry stays unbroken. If $\mu^2 < 0$, then the minimum of the potential is not at zero, but a set of infinite degenerate states at

$$\phi_0^2 = \mu^2 / 2\lambda \tag{3.24}$$

as shown in Fig. 3.1. An appropriate choice of a particular solution, without loss of generality, is the real scalar

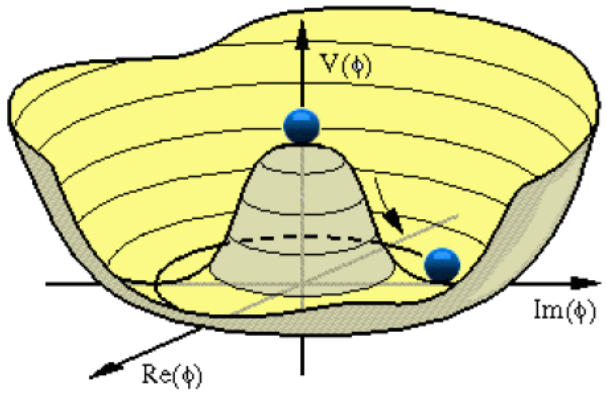


Figure 3.1: Representation of the Higgs potential if $\mu^2 < 0$

$$\phi_0 = -\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \sqrt{\frac{\mu^2}{\lambda}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{3.25}$$

The measured vacuum expected value of this field is of 246 GeV. According to Nambu and Goldstone, rotation around the minimum states of the broken symmetry will give a massless boson. However, the perturbation in the orthogonal direction will give the massive boson. The more general excitation around the ground state

$$\phi_0 = \frac{1}{\sqrt{2}} e^{i\frac{\vec{\xi}(x)\vec{\sigma}}{v}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{3.26}$$

may be rewritten by an appropriate gauge rotation that absorbs the massless and no physical ξ fields, as

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{3.27}$$

For the *Higgs* scalar boson hypercharge Y=1 is chosen so the $U(1)_{em}$ symmetry is not broken, only the weak sector. The BEH field contribution to the Lagrangian is

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi) \tag{3.28}$$

Using Eq. 3.27 and Eq. 3.17 in Eq. 3.28 and then doing a linear combination of the W^i, B fields one gets the $W^\pm$ and Z bosons and a mass term for them.

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2) \tag{3.29}$$

$$Z_\mu = \frac{g}{\sqrt{g^2 + g'^2}} W_\mu^3 - \frac{g'}{\sqrt{g^2 + g'^2}} B_\mu \tag{3.30}$$

$$A_\mu = \frac{g'}{\sqrt{g^2 + g'^2}} W_\mu^3 + \frac{g}{\sqrt{g^2 + g'^2}} B_\mu \tag{3.31}$$

with masses

$$m_W = \frac{gv}{2} \tag{3.32}$$

$$m_Z = \frac{v}{2}\sqrt{g^2 + g'^2} \tag{3.33}$$

$$m_\gamma = 0 \tag{3.34}$$

$$m_H = \sqrt{2}\mu = \sqrt{2\lambda}v \tag{3.35}$$

The Higgs boson mass in Eq. 3.35 is a free parameter depending on the λ factor. Fermions also get their mass thanks to the mass field through terms of Yukawa potential added to the Lagrangian, of the type

$$\mathcal{L}_Y = Y_l \overline{L_L} \phi l_R + Y_d \overline{Q_L} \phi d_R + Y_u i \overline{Q_L} \phi_c u_R \quad + \quad hermitian \;\; conjugates \tag{3.36}$$

where Y_i are the Yukawa couplings, L_L, Q_L are the left doublets for leptons and quarks, l_R, d_R, u_R are the right singlet for leptons, down type quarks and up type quarks, respectively, and ϕ_c is charge-conjugate Higgs doublet where the up quark mass arise from.

For example, for the electron case, using Eq. 3.27 in Eq. 3.36 it results

$$\mathcal{L}_e = -\frac{Y_l}{\sqrt{2}} v(\overline{e_R} e_L + \overline{e_L} e_R) - \frac{Y_l}{\sqrt{2}} h(\overline{e_R} e_L + \overline{e_L} e_R) \tag{3.37}$$

where we have the mass component $m_e = Y_l v/\sqrt{2}$ thanks to the non-zero vacuum expectation value.

3.2.1 Higgs boson production at the LHC

LEP accelerator at CERN (1989-2000), as well as Tevatron at Fermilab (1985-2011), had unsuccessfully searched for the Higgs boson. Exclusion limits were set to certain domains in the mass range [29]. The search and study of this particle is one of the top priorities of the Large Hadron Collider (LHC) program, which will be presented in more detail in chapter 4.
Unfortunately, the Higgs boson mass is a free parameter for the model, so the search strategy must consist in a scan over the mass range domain, and for each mass hypothesis claim certain excess of recorded events or, on the contrary, exclude the hypothesis with a certain confidence level.

The Large Hadron Collider is specifically designed for the Higgs production and detection[30]. There are three modes where the particle can be produced at the LHC, all represented in Fig. 3.2:

- gluon-gluon fusion (gg). This is the dominant production mode at the energies of the LHC. It is created from the interaction of two gluons via a quark loop, usually top.

- vector boson fusion (VBF). It is the second more important case. It is produced by the fusion of 2 vector bosons (W,Z) each of them radiated by a quark-antiquark pair. In consequence, there will be two additional hadronic jets in the final state coming from the radiated quarks.

- associated production. The Higgs will be produced associated with other particles, like a pair of top-antitop quarks (ttH) or a vector boson (WH,ZH). It is very important in studies in the low mass range, (below 160 GeV) because in this mass range the background rate is very high and associated particle can be very helpful in the analysis.

The cross-section depends on the energy of the center of mass and the mass hypothesis. Fig. 3.3 shows the prediction for the two different data taking periods of the LHC, at 7 and 8 TeV of energy in the center of mass, respectively.

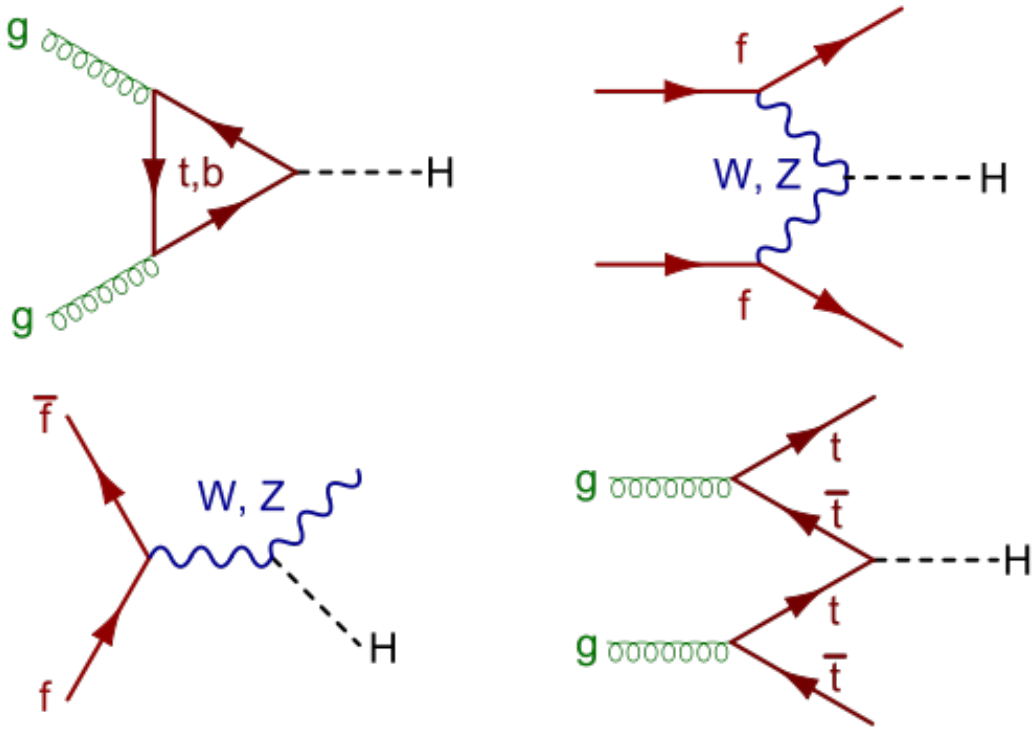


Figure 3.2: Feynman diagrams of the SM Higgs production: gluon fusion (up left), vector boson fusion (up right), vector boson associated (bottom left) and top associated production (bottom right).

3.2.2 Higgs boson decay modes

The Higgs boson is an unstable particle with a very short lifetime. It may decay through different channels, summarized in Fig. 3.4, together with their theoretical uncertainties, coming from missing higher orders in the perturbation theory and from the knowledge of the strong coupling constant and quark masses.

The "branching ratio" or "branching fraction" (BR) is the fraction of the total decays corresponding to a particular final state. As described in section 3.2 the Higgs boson couples both to vector bosons and fermions. At high mass regime, above the invariant mass of twice the W and Z bosons, it will preferentially decay into a pair of WW, mainly, and ZZ. Above the $t\bar{t}$ mass threshold, this decay is also allowed.
At lower masses, the fermionic decays are more likely: principally into a pair of $b\bar{b}$ quarks and into a pair of $\tau\bar{\tau}$ leptons. Higgs decaying into gluon or charm pairs absorbs also an important fraction, but the study of these final states is tremendously difficult in the LHC, especially the case of gluons.
However, the diphoton channel, although much less probable, is an excellent opportunity for its study due to the associated clean signature and the relative low background.
We call background to any physical process different of the process of interest (in this

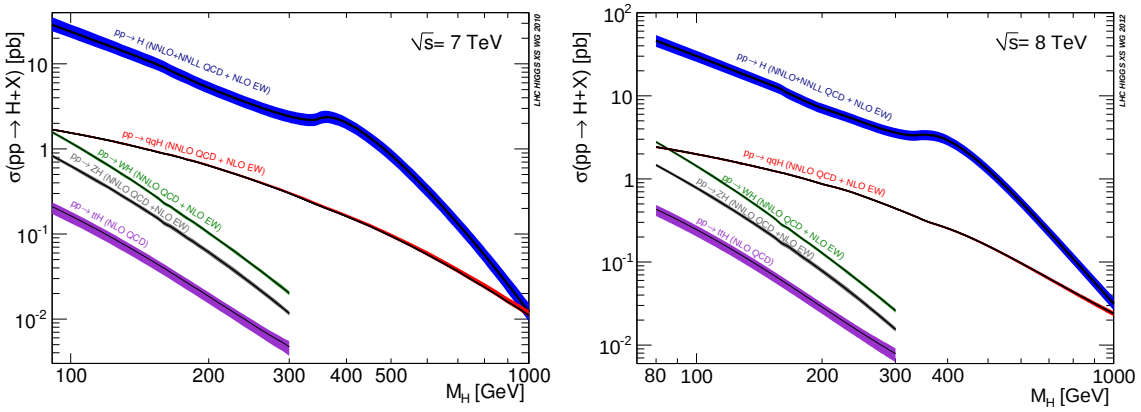


Figure 3.3: SM Higgs cross-section at $\sqrt{s}$=7 TeV (left) and $\sqrt{s}$= 8 TeV (right) at the LHC, as a function of the Higgs mass hypothesis.

case, a Higgs boson production and decay) that will result in the same signature in the detector. Some of them can be distinguished from the *signal process* through different analytical techniques, but many of them are completely indistinguishable, and we call them *irreducible* background.
Figure 3.5 shows the expected cross section times branching ratio ($\sigma(pp \rightarrow H) \times BR(H \rightarrow X)$) for different final states for collisions of $\sqrt{s}$ =8 TeV in the center of mass.

The width of the resonance is highly dependent on the mass hypothesis, being narrow at low masses (∼10 MeV) and rapidly increasing with mass. As the mass hypothesis distances from the electroweak scale (246 GeV) the Breit-Wigner approximation becomes no longer valid to describe it, to the point that in the heavy regime (near 1 TeV) it is not proper to talk about a clear resonance [31]. Figure 3.5 shows the theoretical width as a function of the mass hypothesis. On top of that, quantum interference with SM ZZ and WW production affects the width above the 600 GeV mass hypothesis.

3.3 Results on the Higgs search at the LHC

The fourth of July of 2012 it was announced at CERN the discovery of the last piece of the SM. It meant one of the major achievements in the history of Physics and the biggest success of the LHC program so far: the discovery of a new particle compatible with the SM Higgs boson, carried by both CMS and ATLAS Collaboration independently, with a mass around 125 GeV [18, 19]. Since then, the analysis of the full dataset recorded up to February 2013 has allowed to improve the search of the Higgs boson in all relevant final states, as well as the study of its properties.

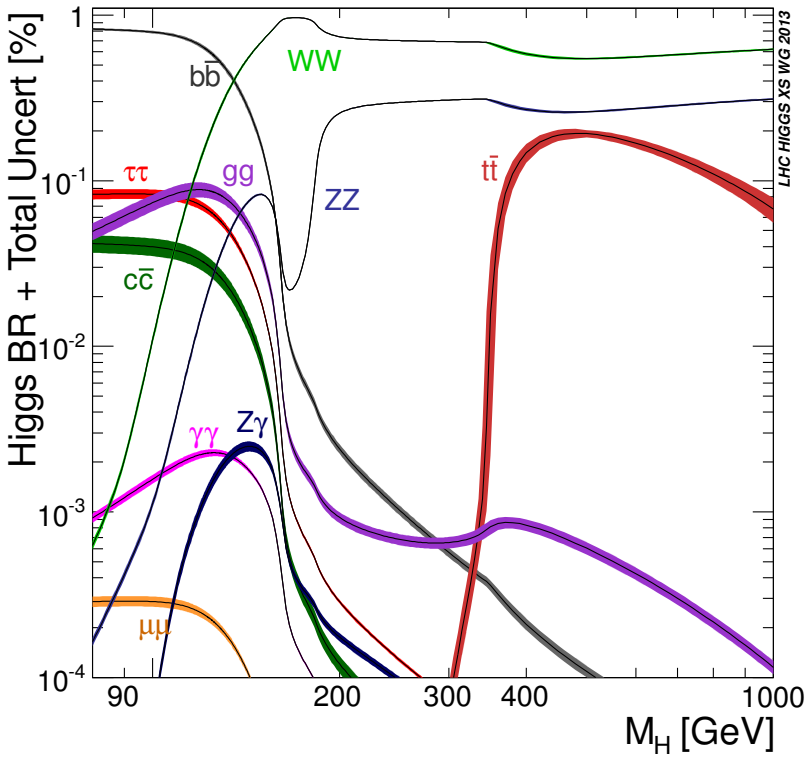


Figure 3.4: Branching ratio of the SM Higgs boson for different decay channels as a function of the mass hypothesis.

3.3.1 Higgs decaying into bosons

The best sensitivity to the Higgs boson has come from analysis of Higgs decaying into bosons (ZZ, $\gamma\gamma$), where it has been observed for the first time. Neither of them is the most likely decay, but this final states possess features (low background processes, high resolution) that allows an optimal study.

Higgs search in the two photon final state

The $H \rightarrow \gamma\gamma$ decay channel provides a clean final-state topology which allows the mass to be reconstructed with high precision, and compensates its low branching fraction (varying between 0.14% and 0.23%). A candidate event is shown in Fig. 3.6 The search is made for a narrow peak in the diphoton invariant mass on a large irreducible smoothly falling background from QCD production of two photons, as shown in Fig. 3.7. An excess of events around 126 GeV is clearly observed [32].

The conclusion is that in all the mass range between 110 and 150 GeV the existence of a SM Higgs boson is excluded at 95% confidence level, except the region between 122.8 and 127.8 GeV where the excess of events is found. The local p-value quantifies the probability for the background to produce a fluctuation as large as the observed one or larger, and displays at m_H=125 GeV a local significance of 3.2 σ (4.2 σ expected),

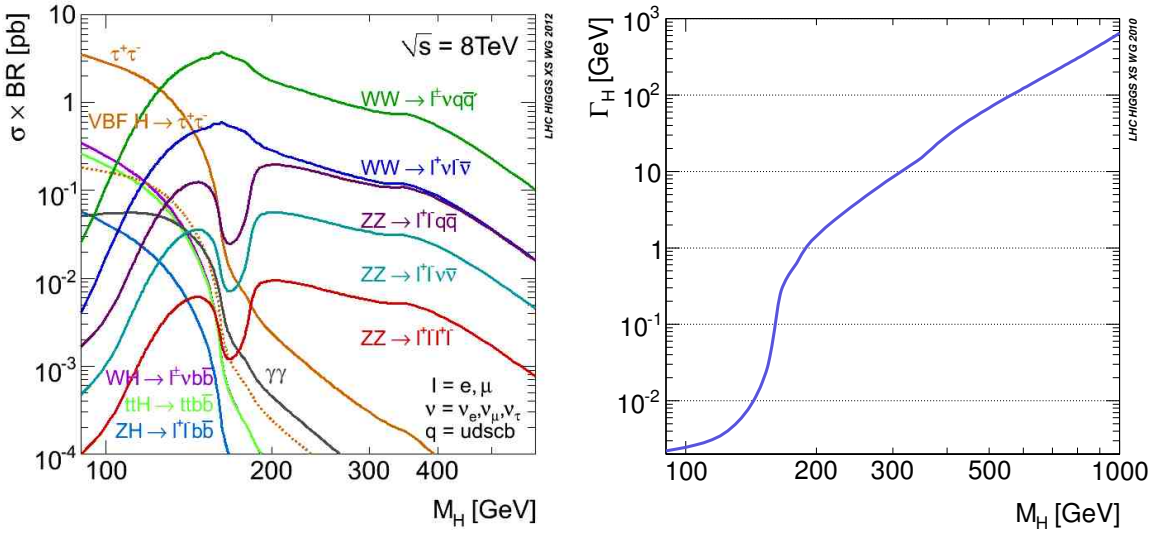


Figure 3.5: Cross section times Branching Ratio of the SM Higgs boson for different decay channels and an energy at the center of mass $\sqrt{s}$=8 TeV, as a function of the mass hypothesis (left). Natural width of the Higgs resonance as a function of the mass hypothesis (right).

as shown in Fig. 3.8. The statistical interpretation of the results will be studied in further detail in following chapters.

ATLAS Collaboration followed a similar approach and found similar conclusions, reaching a 7.4σ of significant excess over a 4.3σ expected [33].

Higgs search in the four lepton final state

Although the $H \to ZZ$ and $H \to WW$ become dominant above ZZ and WW mass thresholds, they still get an important BR at lower masses when one, or the two bosons, are virtual, or "off-shell". The four lepton final state is the most important one for discovery, $H \to ZZ \to l^+l^-l^+l^-$, due to his clean signature and low background, which allows a good mass resolution of the invariant four-object.

In fact, this resonance has been observed with a local significance of 7σ in the ZZ channel alone.

The search of a final state with four leptons (two pairs of electrons or muons, Fig. 3.9) is made for a narrow four-lepton mass peak in the presence of a small continuum background. This allows a large range of study, from 110 to 1000 GeV. The absence of neutrinos and the high resolution of measured momentums allows also a high resolution in the mass. On top of that, kinematical constraints that will be detailed in chapter 8 allow a better discrimination between signal events and background processes. Fig. 3.10 shows the 4-lepton invariant mass distribution, where an obvious resonance is observed at ∼126 GeV [34].

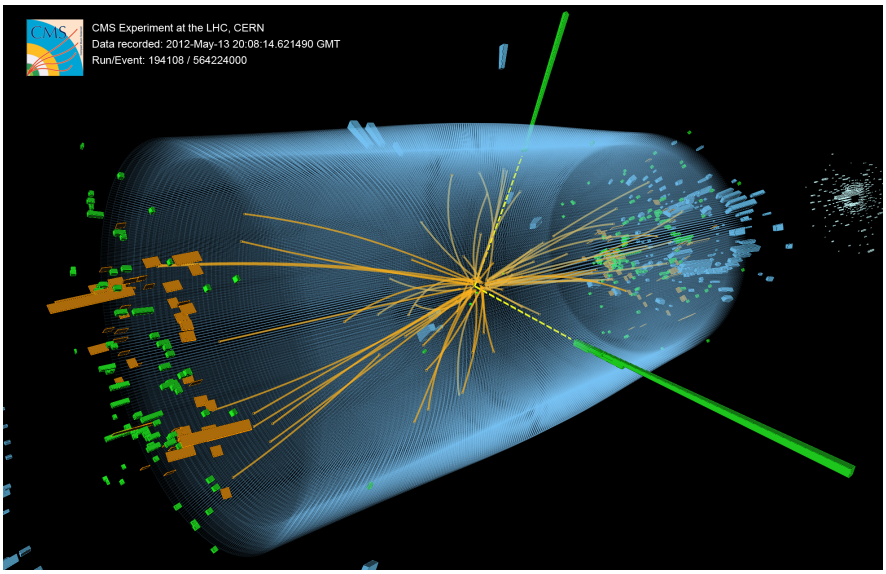


Figure 3.6: Event display of a candidate of a Higgs boson decaying into two photons, recorded by the CMS Collaboration

Higgs in the WW* final state

Although the H→ WW^* shows the highest BR from 160 GeV, it is possible to extend the search to 120 GeV when the two W decay leptonically, through different analysis techniques. Unfortunately, the mass resolution is degraded due to presence of undetectable neutrinos in the final state, so a clear peak is not possible to be seen, but an excess of events around the mass of the boson, which has been observed with a local significance of 3.8σ in the case of ATLAS and 4.0σ in the case of CMS [35].

3.3.2 Higgs decaying into fermions

Both CMS and ATLAS have shown strong evidence for the decay of Higgs bosons into fermions. The measured decays are the pairs of bottom ($b\bar{b}$) quarks and pairs of $\tau\bar{\tau}$ leptons, which are the heaviest fermions it could decay with this mass. The b-quark channel is the most common of all Higgs decays at this mass but background processes make it very difficult to extract a signal due to a large background rate. On the other hand, the tau channel has a cleaner signature. Overall, CMS claims a combined significance of 4.0σ, which compares well with ATLAS result of 4.1σ. Left plot of Fig. 3.11 shows the observed signal strength compared to the SM expectation for both analysis [36, 37, 38, 39].

3.3.3 Properties

The general conclusion from the results is the impressive consistency with SM. Strong deviations from it are already excluded, although current uncertainties allow

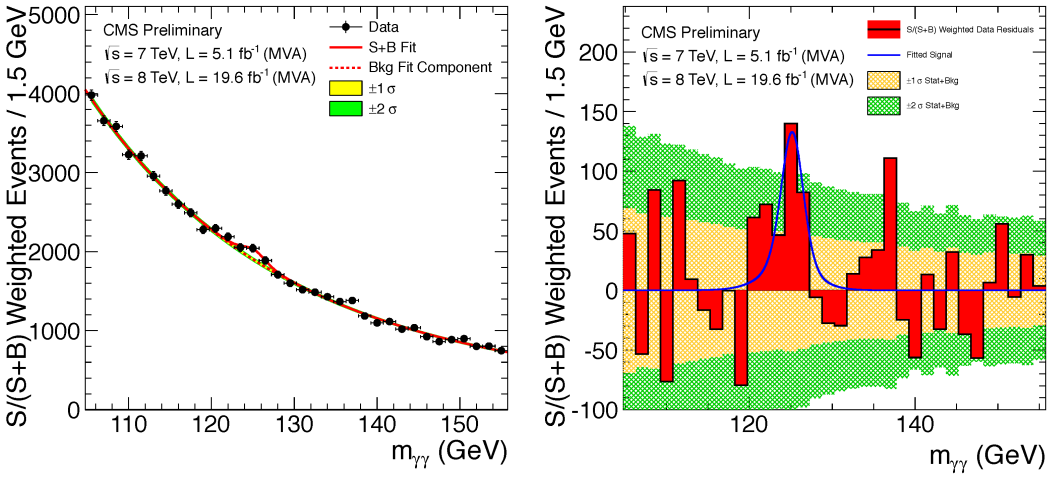


Figure 3.7: Diphoton invariant mass distribution (left) and background subtracted diphoton invariant mass distribution (right), with each event weighted by the S/(S+B) value of its category, for the mass-fit-MVA analysis on the 7 and 8 TeV data sets combined by CMS.

more exotic scenarios to be possible [40].

The Higgs mass is a very important result, as it is the last parameter of the SM, and consequently of our knowledge of Nature, to be measured. The value obtained for the mass combining the measurements from all channels is $\mathbf{m_H = 125.5 \pm 0.2\ (stat.)\ +0.5\ -0.6\ (syst.)\ GeV}$ in the case of ATLAS[41] and $\mathbf{m_H = 125.7 \pm 0.3\ (stat.) \pm 0.3\ (syst.)\ GeV}$ in the case of CMS [40].

The signal strength, which accounts for the cross section referred to the SM case, and including the mentioned decay channels, is slightly different for ATLAS and CMS, but consistent with the SM considering the uncertainties.

- ATLAS ($\gamma\gamma$,WW*,ZZ*,bb,$\tau\tau$): μ= 1.23±0.18
- CMS ($\gamma\gamma$,WW*,ZZ*,bb,$\tau\tau$): μ= 0.80±0.14

One important thing to study is the coupling strength to the other particles, bosons and fermions, as a clear hint of deviations from the SM. The right plot of Fig. 3.11 shows the couplings, referred to the SM expectation, for the different processes.

Finally, the study of spin and parity of this new particle is crucial to determine its nature and the role of the scalar sector in Nature. Spin-1 hypothesis is discarded by the observation of H $\rightarrow \gamma\gamma$, so the main analyses have tested the hypothesis of 0^- Vs 0^+ and of 2^+ Vs 0^+. The 2^+ hypothesis, for example, would be produced by a graviton in certain models. Both hypothesis have been excluded by CMS (and also ATLAS) at more than 99% CL, as shown in Fig. 3.12 [42, 43, 34, 44].

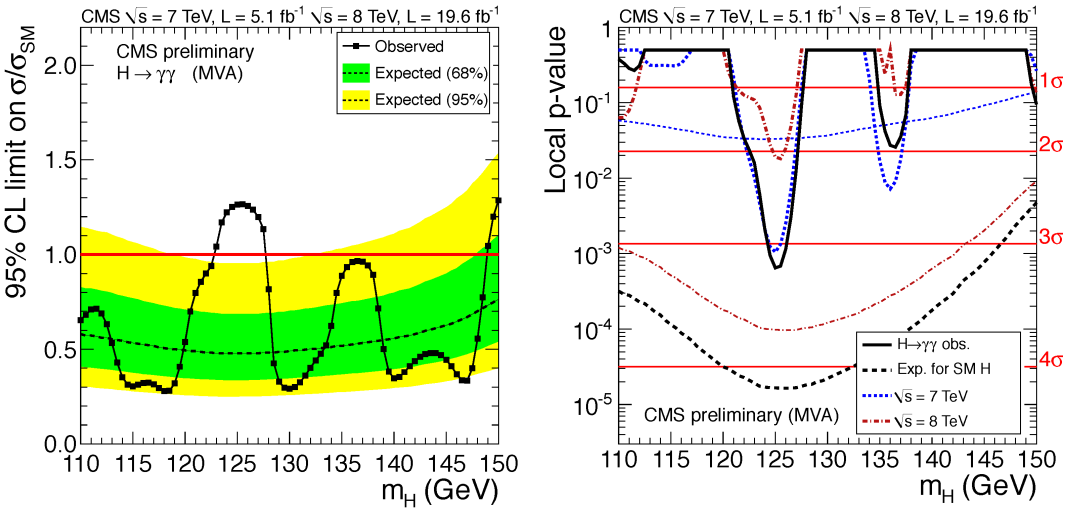


Figure 3.8: The 95% CL limit on σ/σ_{SM} (left) and observed local p-values as a function of mH (right) for a Higgs boson decaying to two photons, obtained in the mass-fit-MVA analysis for the 7 and 8 TeV data sets combined by CMS.

3.4 Mass Mechanism Beyond the Standard Model

The Standard Model of Particle Physics has been successfully tested over the last decades, and it is the best explanation the human being has of Nature. However, several phenomena has been observed inconsistent with the SM. It is known that it cannot be the last theory, but an approximation to some more general theory that also explains some of the missing problems of the SM:

- The Standard Model does not explain why only 3 generation of particles exist. No fourth generation has been observed so far, but it does not mean it does not exist.

- There is no quantum theory for gravitation. All attempts of combining the General Relativity with a Quantum Field theory have failed. To be fair, the source of the gravity is not clear yet. The big problem to correctly understand gravity comes from the fact it becomes negligible at the scale of the laboratory compared to the other interactions, so no experiments can be done as with the latter.

- In the SM, neutrinos are massless. However, the observed oscillation of neutrino flavour implies nonzero masses for them, regardless how much small they were. On top of that, it is still an open question whether neutrinos are Dirac or Majorana spinors, which would imply New Physics beyond SM.

- The SM cannot explain the asymmetry matter-antimatter observed in the Universe. The known weak CP violation included in the SM is not big enough to explain it.

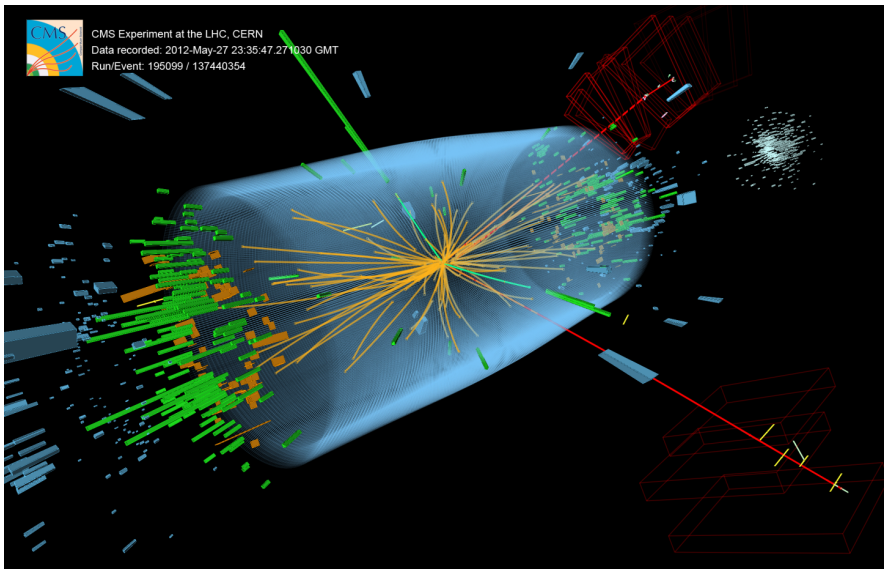


Figure 3.9: Event display of a candidate of a Higgs boson decaying into 2 muons and 2 electrons, recorded by the CMS Collaboration

- Dark Matter cannot be justified with the Standard Model (nor the General Relativity). If it is caused by some particle, it is not known. Many models have been proposed and many of them are being tested in the LHC.

- One topic that disturbs the scientific community is the *hierarchy problem*; our inability to explain why couplings of different interactions are so different (in particular gravity and weak force). This is related with the fine tuning fundamental parameters in Nature seem to have, contrary to the idea of *naturalness*.

Many attempts to develop theories beyond SM have been presented in the last decades. Supersymmetry [45, 46] introduces a new symmetry between fermionic and bosonic fields, implying the existence of a superpartner for every known fundamental particle: every fermion is paired to a s-fermion, which is a boson, and vice versa. It solves hierarchy problem, allows unification of the three interactions at the grand unification theory scale and provides possible candidates for the dark matter. However, if the symmetry was real, sparticles would have the same mass as particles, so this Supersymmetry has to be somehow broken.

Another approaches include Extra-dimensions, Little Higgs, Technicolor, String theory, Grand Unified Theory (GUT), etc. These models use to include a large number of extra free parameters and some of them have been tested in past accelerators and in the present LHC.

In what concerns the mass generation of fundamental particles, and despite the discovery of a Higgs boson, it is still early to claim that the SM mass mechanism is completely correct. This mechanism is the simplest one, but other models could

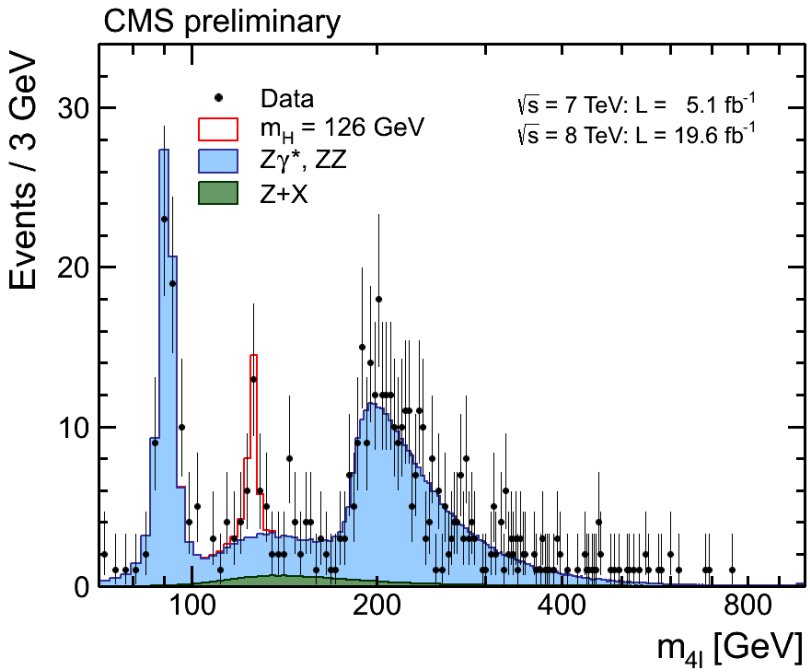


Figure 3.10: 4-lepton invariant mass distribution obtained by CMS. It includes 4-muon, 4-electron and 2-muon plus 2-electron final states.

be consistent with the data given its current uncertainties. Some of them predict additional resonances, which would be a hint of them [47].

For example, the most simple supersymmetric case, the Minimal Supersymmetric Standard Model (MSSM), requires at least a second fundamental Higgs doublet to preserve the cancellation of gauge anomalies. In GUT, a extended scalar sector is needed to break whatever Lie group of a unified theory into the $SU(3) \otimes SU(2) \otimes U(1)$. In the simplest extended case, only 2 complex doublets are considered:

$$\phi_i = \begin{pmatrix} \phi_i^+ \\ \phi_i^- \end{pmatrix} \tag{3.38}$$

The Lagrangian would be a extended version of the SM case

$$\mathcal{L}_H = \sum_i (D_\mu \phi_i)^\dagger (D_\mu \phi_i) - V(\phi_i^\dagger \phi_i) \tag{3.39}$$

with a potential that, in the most general case considering quadratic and quartic terms

$$V(\phi_1, \phi_2) = \sum_{i,j=1,2} \mu_{ij}^2 \phi_i^\dagger \phi_j + \sum_{i,j,k,l=1,2} \lambda_{ijkl} (\phi_i^\dagger \phi_j)(\phi_k^\dagger \phi_l) \tag{3.40}$$

would have 14 free parameters, instead of the 2 of the SM case. Depending on the specific model and the simplifications considered, this number can vary. As a consequence of Eq. 3.38, there will be eight degrees of freedom. If, as in the SM, 3 of

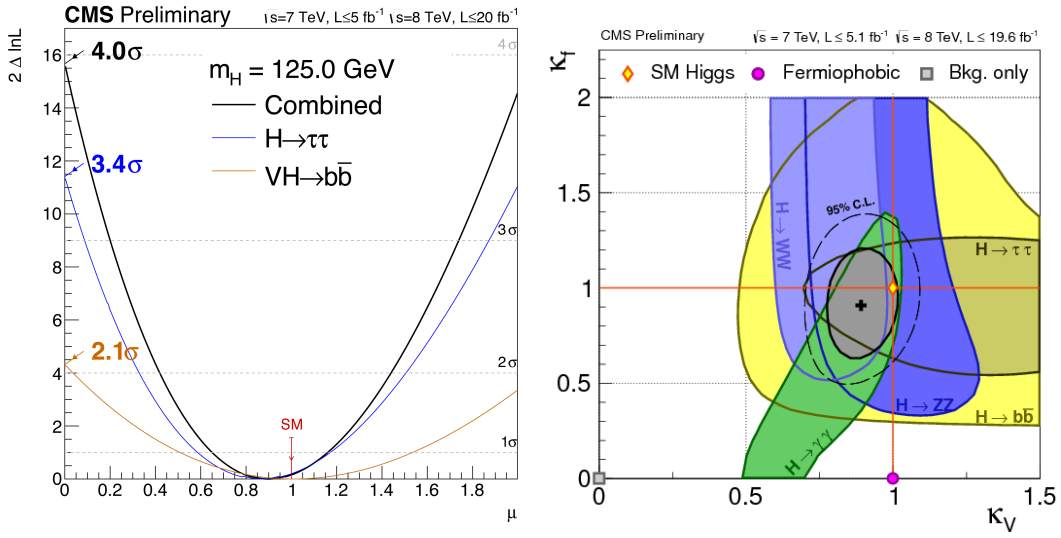


Figure 3.11: Left: The most likely value for the ratio of the signal strength relative to the Standard Model expectation (μ). Right: 2-D scan of Higgs boson couplings, referred to the SM expectation, for the different processes, as measured in CMS.κ_v refers to couplings to bosons, and κ_f to couplings to fermions.

the fields make the W,Z bosons massive, in this case five physical Higgs bosons should exist: three neutral ones h_1, h_2, h_3 and two charged $H^{\pm}$.

Depending on the particular model and fine tuning, an additional Higgs-like (i.e., with the same signature than the SM Higgs boson) resonance is predicted with a heavier mass in the range up to 1 TeV. Consequently, dedicated effort must be put in searching of these signatures.

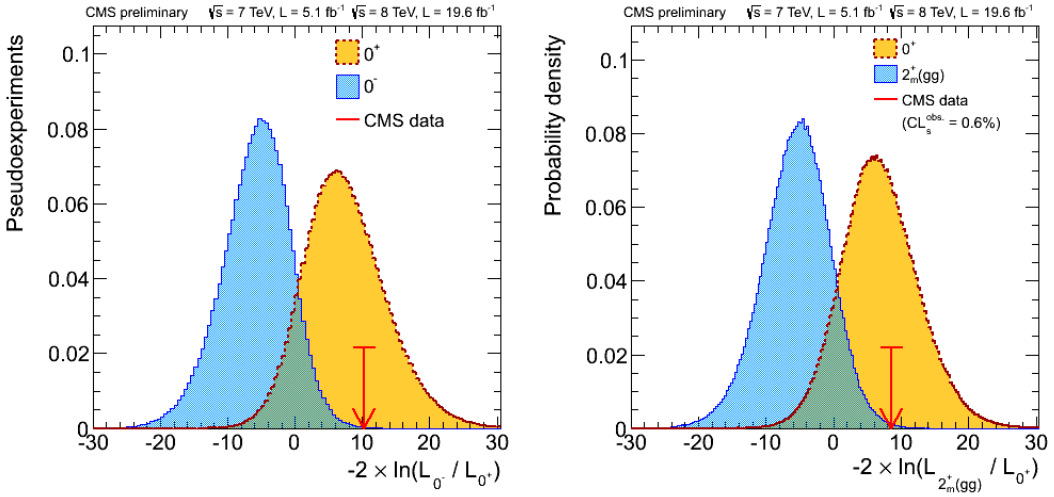


Figure 3.12: Test of the 0^- (left) and 2^+ (right) hypothesis against the scalar 0^+ one, in CMS. Both plots show that the new particle discovered is compatible with a scalar boson.

Chapter 4

The Large Hadron Collider

The Large Hadron Collider [48, 49, 50, 51] is the largest and most ambitious accelerator ever built up to now. Almost 27 kilometer long, it is placed in a circular tunnel between 50 and 150 meters underground, and provides ultra-high energy collisions to the experiments in the LHC program since 2009. This chapter presents a brief description of the LHC.

4.1 CERN and the LHC

The European Organization for Nuclear Research (CERN), founded in 1954 [52], has been a reference in Fundamental Research for the last half century. It is based on four purposes:

- A pure scientific and fundamental research in Nuclear and High-Energy Physics, studying what is Universe made of and how it works.
- Development of related technologies with applications in other fields (medical, etc).
- The dissemination of scientific knowledge: Training of young scientist and engineers and promotion of science to the general public and students.
- Promotion of international collaboration, involving 11 000 scientists, representing 641 universities all around the world, by the year 2013.

CERN accounts for a chain of several accelerators of different energy (Fig. 4.1), being the biggest and newest the LHC. The others are used to provide particle beams different experiments but also to accelerate them before entering the LHC for the final acceleration step.

The LHC is conceived as a discovery machine, in contrast to other accelerators focused in the accuracy of the parameters measured. It produces high energy collisions in four points of the ring, where independent experiments detect and record the

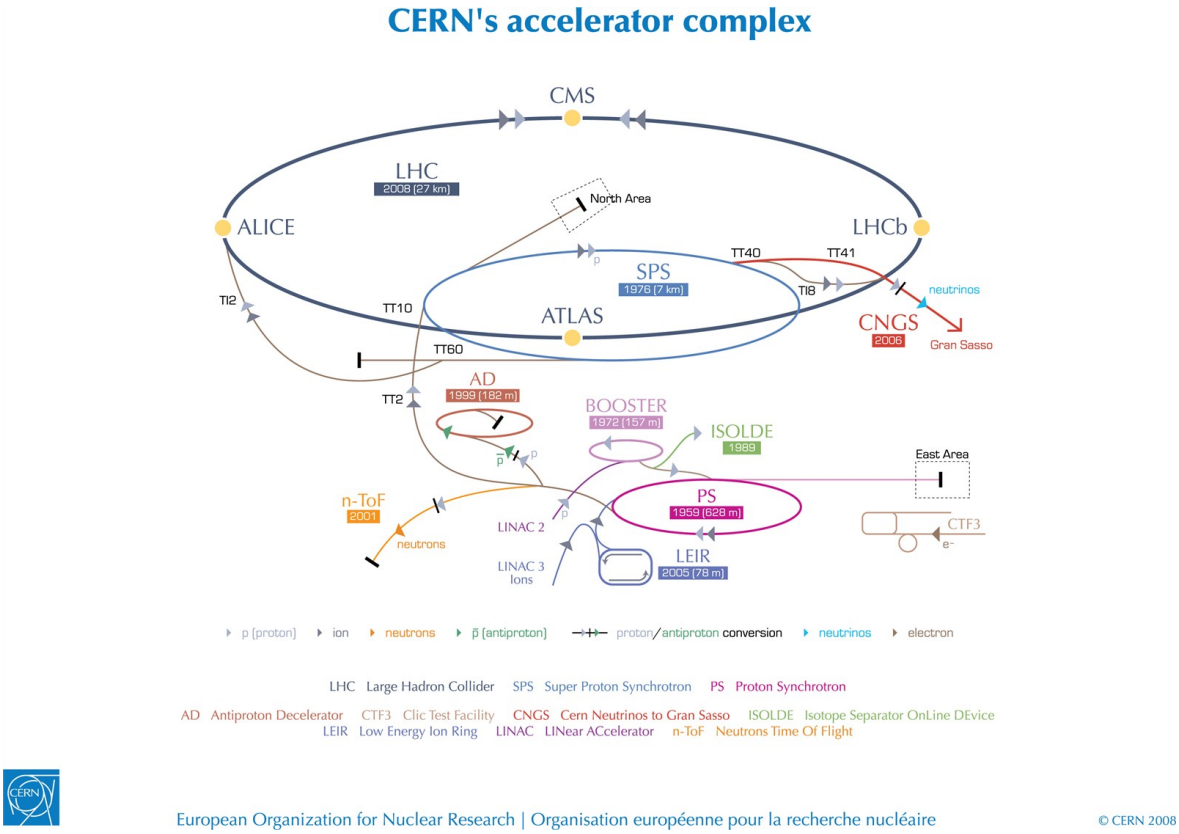


Figure 4.1: The CERN accelerator complex.

products of those collisions, as shown in Fig. 4.2. LHCb [53] is designed to study CP violation, and the matter-antimatter asymmetry, through processes related with the b-quark. ALICE [54] is a dedicated heavy-ion detector which benefits of the lead collisions to study strongly interacting matter at extreme energy densities where quark-gluon plasma is expected. ATLAS [55] and CMS[56, 57] are the biggest collaborations, and their detectors, called *multipurpose*, are designed to perform all type of analysis, from Higgs searches, New Physics searches (Supersymmetry, extra-dimensions, exotic theories, etc), precise SM measurements or quark-gluon plasma studies.
On top of that, TOTEM [58, 59] is a smaller detector placed next to CMS aiming total elastic and diffractive cross-section measurements. LHCf [60] is placed near ATLAS and measures particles with low angle with respect to the beam direction, to simulate cosmic rays. Finally, MOEDAL [61] searches directly for magnetic monopoles.

The present thesis has been realized inside the CMS Collaboration.

4.2 The LHC chain

The LHC has two operational modes, where the beams are composed of protons or lead nuclei, respectively. The latter are produced to understand matter at high energy densities and study the quark-gluon plasma. There has been also a small period of p-Pb collisions, to improve the understanding of cold nuclear matter and cosmic rays, and complement the Pb-Pb results. However, most of the Physic aims are studied in pp collisions, less complex, which cover also most of the time of data taking.
In this scenario, hydrogen gas atoms are extracted from a gas bottle and turned into plasma with electric fields, to separate protons and electrons. Protons are derived to the Linear Accelerator 2 (LINAC 2), which accelerates the protons to 50 MeV. The *Proton Synchrotron Booster* (SPB) accelerates the protons to 1.4 GeV followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV in its 628 metres. Protons are then sent to the Super Proton Synchrotron (SPS), 6.9 km long, where they reach 450 GeV, moment they are injected in the LHC. The PS and the SPS have been the world's highest energetic accelerator in the time of their running beginning.

In the lead case, the starting point of ion's journey is the Linear Accelerator 3 (LINAC 3), where Pb^{+54} is produced. They continue accelerating in the Low Energy Ion Ring (LEIR) up to 72 MeV, to be dropped in the PS, where the ultimate Pb^{+82} is obtained. From the SPS the chain is the same as for the protons. In the following, only proton-proton mode (pp) will be considered concerning the beam discussion.

4.3 The LHC design

The LHC was approved on 1994 and construction ended in 2007. Collisions started in November 2009 after commissioning process. It is placed in a tunnel 26.7 km long crossing the border between Switzerland and France, and divided in eight sectors. Acceleration is achieved in superconductive radiofrequency cavities (RF) which store a 400 MHz stationary electric field of 2 MV of potential energy. RF's are also responsible of keeping the particles in small packets called bunches. There are a total of 16 RF's in the LHC, 8 for each beam. All are located in only one place along the ring, represented in Fig. 4.2.

The 26.7 km of the ring are designed to conduct and focus the beams. This is reached with superconductive magnets. 1232 dipoles, as the one pictured in Fig. 4.3, bend the trajectory of the beam to keep it in the center of the pipe. 392 quadrupoles optically focus the beam, which naturally tends to spread due to repulsive electromagnetic forces among the charged particles, maximizing the probability of collisions. Sextupoles, octopoles, decapoles and dodecapoles introduce minor corrections to the beam.

Two beams circulate in opposite directions. As both beams are same sign charged, they must be in separate pipes with opposite magnetic fields. This is obtained with

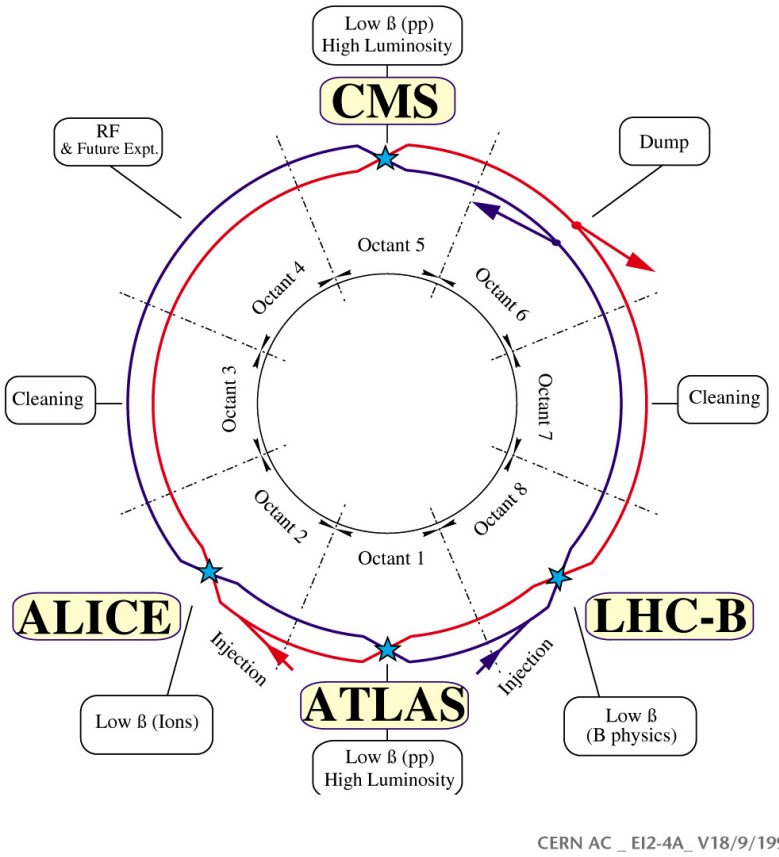


Figure 4.2: The LHC layout with the location of the main experiments..

a specific magnet design. Magnets are made of an alloy of Nb-Ti of wire braid which stands 13 kA to produce a nominal magnetic field of B=8.3 T.

To keep the magnet in superconductive state, the central structure of the accelerator, called *cold mass* is cold to 1.9 K with superfluid helium.

Pipes where particles travel have 6 cm of diameter and are kept at 5 K. Ultra high vacuum of $\sim 10^{-9}$ Pa is produced inside.

Some of the most relevant parameters of the machine and their nominal values are listed in Table 4.1.

4.4 Energy and luminosity

The LHC has been designed to produce collisions with a nominal energy of $\sqrt{s}$=14 TeV in center of mass (c.o.m) frame. However this value has not been reached yet. The LHC started on November 2009 with 1.18 TeV per beam (2.36 TeV in the c.o.m). On March 30^{th} 2010 the first collisions at $\sqrt{s}$=7 TeV took place, which was maintained through 2010 and 2011. During 2012 up to the end of the data taking

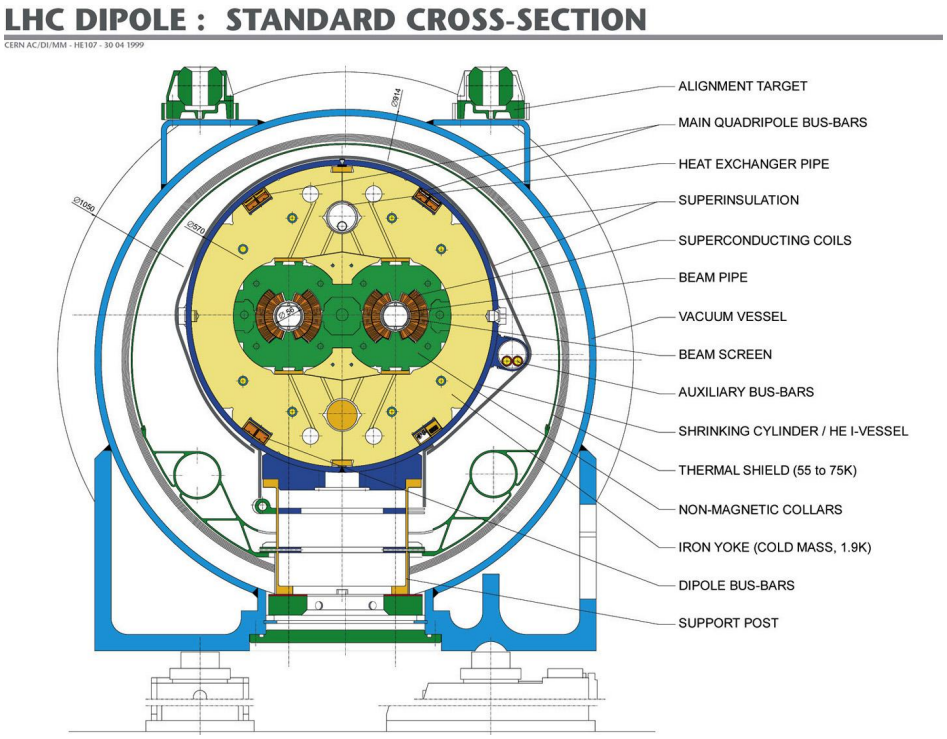


Figure 4.3: A LHC dipole schematic drawing.

period, the energy was raised to $\sqrt{s}$=8 TeV. From February 2013, the LHC is in *Long Shutdown 1*, a period for improvements in the accelerator and in the detectors. Collisions will be restarted, presumably in 2015, with an energy closer to the nominal value. This thesis will cover the analysis $H \to ZZ \to l^+l^-q\bar{q}$ performed in the two runs at 7 and 8 TeV, with more attention in the latter.

The number of events of a certain process depends on the specific cross section of the process and on the luminosity, where the production rate is given by

$$R = \mathcal{L} \cdot \sigma_{pp \to X} \tag{4.1}$$

being $\mathcal{L}$ the instant luminosity delivered by the accelerator and $\sigma_{pp \to X}$ the cross section of the $pp \to X$ process under study.

The instant luminosity depends on the LHC performance and is expressed in terms of several design parameters:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{4.2}$$

where

- N_b is the number of particles per bunch,
- n_b the number of bunches per beam,

Design Parameters	
Circumference Length	26659 m
Depth	50 - 175 m
Total number of magnets	9600
Number of main Dipoles	1232
Time between collisions	25 ns
Bunch Crossing Rate f_{rev}	40.08 MHz
Temperature	1.9 K (-271.3C)
Injection energy	450 GeV
Dipole field	8.33 T
Nominal proton energy	7 TeV
Centre-of-mass energy $\sqrt{s}$	14 TeV
Number of bunches per proton beam	2808
Protons per bunch	1.1×10^{11}
Bunch spacing	25 ns
ϵ_n Norm. transverse emittance	3.75 μm
Transverse beam size at IP5	16.7 μm
$\beta^\star$ IP5 beta value	0.55 m
Crossing angle	285 μrad
E_{stored} Stored energy	362 MJ
Design instantaneous luminosity	10^{34} cm^{-2}s^{-1}
Average number of collisions per crossing	20

Table 4.1: Design parameters of the Large Hadron Collider.

- f_{rev} the revolution frequency,
- γ_r the relativistic gamma factor,
- ϵ_n is the normalised transverse beam emittance, defined as the product of the RMS of the particle position distribution times the RMS of the particle momenta distribution, σ'. It gives an idea of the spacial and the momentum dispersion of the beam. This parameter is constant for all the beam life and the goal of the injection procedure is to introduce in the collider a beam with the lowest beam emittance.
- β^*, is the beta function at the collision point, defined as the ratio σ/σ'. It is reduced along the ring length, with a dedicated magnetic optics which squeezes the beams in the Interaction Point (IP). Note that a lower β^* means a higher spread in the particle momenta.
- F is the geometric luminosity reduction factor due to the crossing angle at the interaction point.

The small production rate of the interesting processes makes necessary large samples of recorded data, so one of the goals of the LHC was to increase the luminosity, up to close the nominal value $\mathcal{L} \sim 7.7 \cdot 10^{33} cm^2 \cdot s^{-1}$ achieved during 2012. The peak luminosity of the LHC is represented as a function of the time in Fig. 4.4. The maximum number of bunches per beam was 1380, with a 50 ns of separation, which is two times longer than the nominal value. The number of particles per bunch and the separation among them define the *number of pile-up*, or number of simultaneously interactions in the detector, which is one of the main problems to deal with. These values were established as a compromise among the capacity of the accelerator to deliver collisions and the ability of the detectors to correctly handle them.

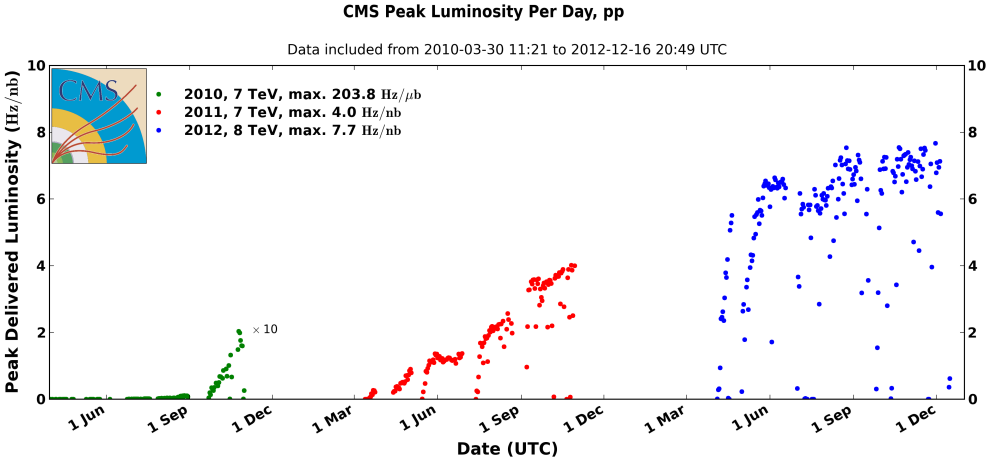


Figure 4.4: Peak luminosity versus day delivered to CMS during stable beams and for pp collisions. This is shown for 2010 (green), 2011 (red) and 2012 (blue) data-taking.

In summary, the LHC showed an extraordinary performance from its beginning, improving the beam parameters (luminosity, energy...) over the years and granting the detectors to fulfill their physics objectives.

Chapter 5

The CMS detector

The Compact Muon Solenoid [56, 57] detector is one of the four main experiments at the LHC. It is placed in a 100 m underground cavern in the point 5 of the LHC, near Cessy (Gex, France). It is a general purpose experiment, aiming to study in detail both the parameters of the SM, and any hint of New Physics beyond the Standard Model. This chapter summarizes the characteristics of the CMS detector.

The CMS experiment belongs to an international collaboration of more than 4000 people from more than 150 institutes around the world. Construction of the detector started in 1998 and finished 10 years later. CMS is called "general purpose experiment" as it is conceived to study all topics of interest in range of energy of the LHC [62].

In particular, the main goal of its design was the discovery of the Higgs boson, already discussed in chapter 3, and particles beyond SM, like sparticles. Other theories more exotic than Supersymmetry include large extra dimensions, heavy vector bosons W' and Z', heavy stable charge particles, fourth generation, exotic resonances ... in general, any model with a measurable signature by the detector.

Furthermore, there is an intense program in the measurements of the SM parameters. In particular, the third generation of quarks is an intense field, specifically about the measurements involving the top quark: cross section, mass, charge asymmetry, polarization, spin correlations, etc. And more generally, all production cross sections, multiboson and multijet production, strong coupling constant and multiple SM parameters.As a example, a summary of the SM cross sections measured in CMS is shown in Fig. 5.1.

Several requirements are necessary in order to achieve all this: First, a good muon detection. Many of the topics of interest include muons in their final state, so an excellent muon identification and resolution is desired, as the name of the detector suggests. Second, the best possible charged particle momentum resolution, which depends on a strong magnetic field to bend the tracks. A good tracking efficiency and resolution is also essential to correctly reconstruct all vortices's in a bunch crossing coming from simultaneously independent collisions (pile-up). Next, a good hermeticity and to cover all the solid angle and provide an accurate estimation of the missing energy. A good calorimetry and jet energy determination is very important. Finally,

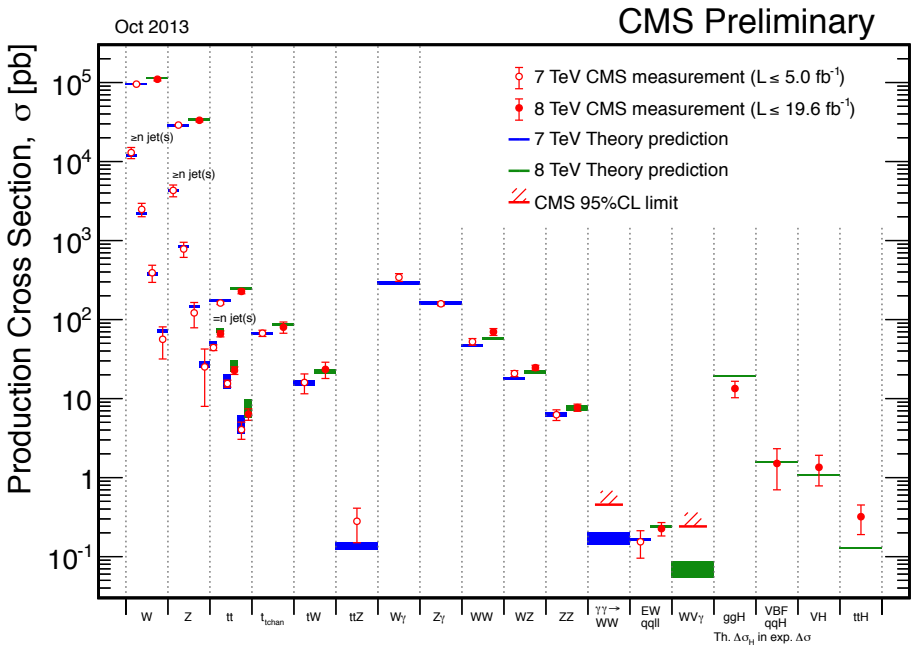


Figure 5.1: Summary of inclusive production cross sections measured in collisions at $\sqrt{s}$=7 and 8 TeV.

a trigger system able to deal with the large luminosity of the LHC and reduce the readout rate to a manageable quantity.

The CMS detector is a solid horizontal cylinder of 16 m high and 21.7 m long, with a total weight of 14000 tons. It is divided in a central region and two endcaps which close it at both sides, as shown in Fig. 5.2.

It tries to be as hermetic as possible, and only the region of the beampipe is not instrumented for detecting particles. It is common to use a magnitude called *pseudorapidity*, which is related with the polar angle and with the longitudinal momentum of a particle

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) = \frac{1}{2} ln \left(\frac{\mathbf{p} + \mathbf{p_L}}{\mathbf{p} - \mathbf{p_L}}\right) \tag{5.1}$$

Different subdetectors cover different pseudorapidity ranges. Acceptance of CMS calorimeters, key in the missing energy reconstruction, extend over a range of $|\eta| < 5.1$.

The key part of CMS is a superconducting solenoid providing a strong magnetic field in the direction of the beam. This magnetic field bends the trajectory of charged particles, like muons, and makes possible the measurement of its momentum. The detector is segmented in concentric layers of various detectors, with different purposes, depending on the interaction of different particles with the materials. Figure 5.3 illustrates in which layer each of the particles are detected in CMS. The most inner part is a silicon tracker detector, which detects charged particles, and measures with excellent precision their trajectory, and hence, their momentum. The second inner

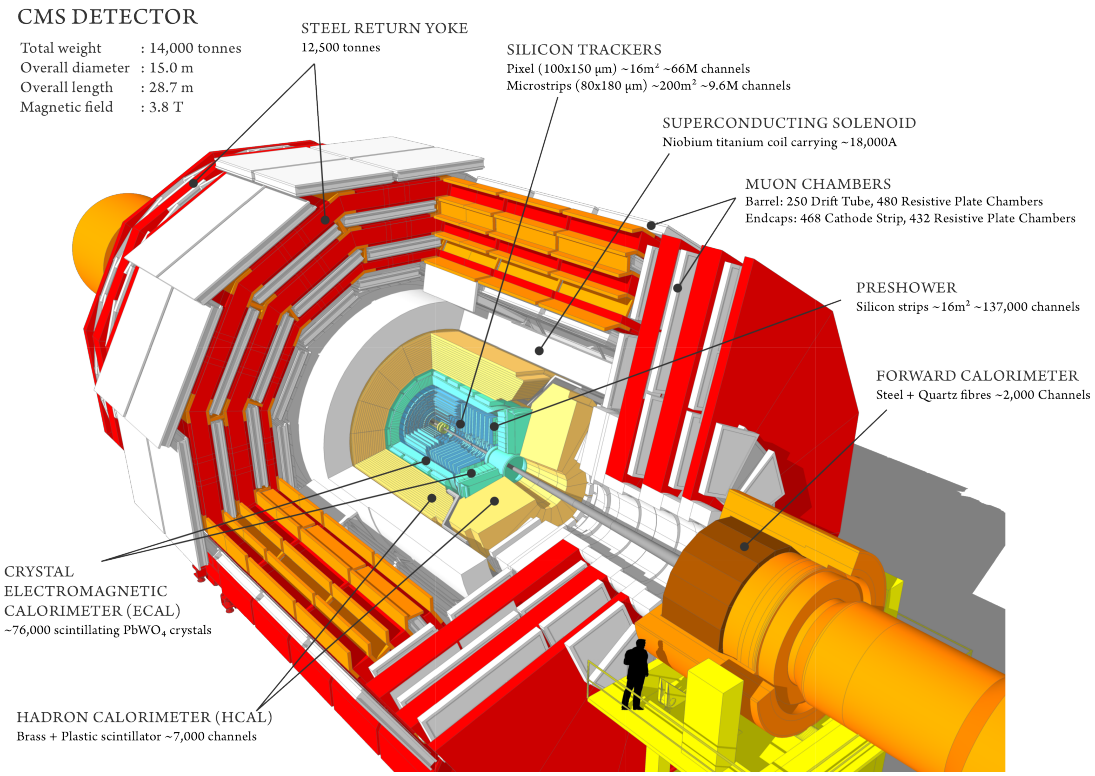


Figure 5.2: Drawing of the CMS detector and its several subdetectors.

layer is an electromagnetic calorimeter, which collects electrons and photons, creating an electric signal proportional to the energy of the particle. Hadrons and mesons cross the electromagnetic calorimeter and are fully absorbed in the hadronic calorimeter. Lastly, outside the solenoid, the muon system identify muons and measures, with the tracker detector, its momentum with great accuracy.
In the following sections, the different parts are described in more detail.

5.1 Magnet

CMS magnet is a superconducting solenoid of 3.8 T which stores 2.6 GJ at full current, which makes it the world' s strongest magnet of its kind. It is 12.5 m long and with an inner radius of 4.9 m. Silicon tracker and the two calorimeters in the barrel are placed inside. It is made of four-layers winding of Nb-Ti conductor stabilised and reinforced by aluminium.

An iron yoke serves as structural support of the whole detector, and for the return of the magnetic flux. Approximately two thirds of the magnetic flux return through the barrel yoke, so the magnetic field is negligible inside the muon chambers, as desired. The remaining third of the total flux returns outside the iron yoke.

In total, the magnet system weights more than 12000 tons.

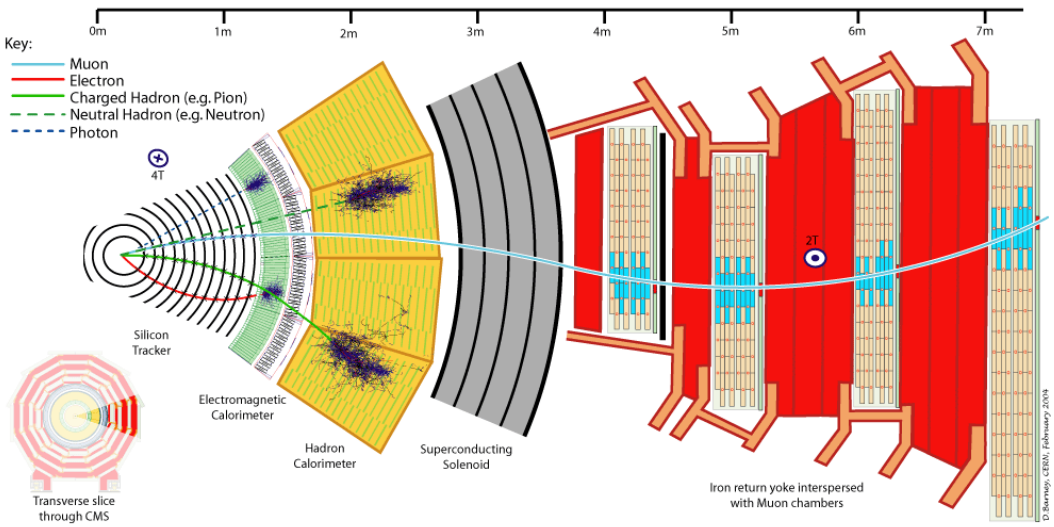


Figure 5.3: Slice of the CMS detector and the area of detection of different types of particles.

5.2 Tracking system

The most inner part of the detector is a tracker made of silicon [63]. It is responsible of detecting charged particles and measure their trajectories (tracks), so their momentum is estimated and the primary vertex of the collision can be reconstructed.

With the increase of luminosity by the LHC, the number of interactions per bunch crossing increases and an efficient and high-accurate tracker is needed to identify and separate the different interaction points, and calculate the impact parameter[1] of different tracks that is used in different algorithms, like b-quark identification (commonly known as *b-tagging*). The number of pile-up has raised to an average value of 21, as observed in Fig. 5.4.

The subdetector is composed of several layers of silicon. When a charged particle crosses the semiconductor volume under an electric field, it promotes some electrons to the conduction band creating an electric signal. Several concentric layers allow to reconstruct the trajectory and create an accurate 3-D picture of the event.

Silicon tracker in CMS is very ambitious, with an efficiency higher than 95% for isolated tracks and 90% for tracks within jets. It covers all ϕ angle and a pseudorapidity range $|\eta|$ <2.5. The primary vertex resolution is 10-20 μm, whereas the overall CMS expected resolution in momentum for charged particles varies from $\Delta p_T/p_T = 1-2\%$ for p_T ~100 GeV up to $\Delta p_T/p_T = 10\%$ in the TeV scale.

It is divided in two parts, detailed in Fig. 5.5. The first part, made of 3 layers of pixels of only 4.4, 7.3 and 10.2 cm radii. Each pixel is 125 x 125 μm^2, with a spatial resolution of 10-15 μm, making a total of 66 million readout channels. The

[1]Distance perpendicular to the beam axis

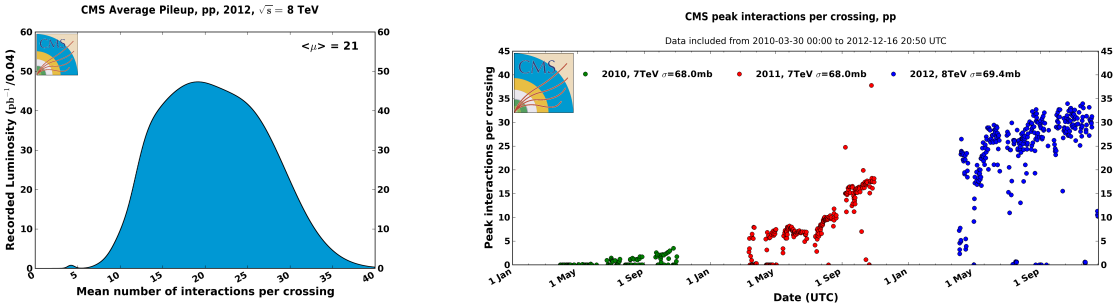


Figure 5.4: Left: Mean number of interactions per bunch crossing in 2012. Right: Peak Interactions per crossing versus time for pp collisions. Each color represents a year of data taking: 2010 (green), 2011 (red) and 2012 (blue).

second is made of 10 layers of microstrips with 320 to 550 μm pitch (depending on the layer/location), placed up to 130 cm distance of the IP, making a total of 80,000 chips and 10 million readout channels. The Tracker Silicon Strip detector is divided in four inner barrel (TIB) layers assembled in shells with two inner endcaps (TID), each composed of three small discs. The outer barrel (TOB) consists of six concentric layers. Finally two endcaps (TEC) close off the tracker. The strip resolution depends on the strip size and ranges about 20 to 50 μm. All together, with more than 200 m^2 of silicon sensors, the CMS silicon tracker is the biggest ever built, needing a cooling system to be kept safe at -20 °C.

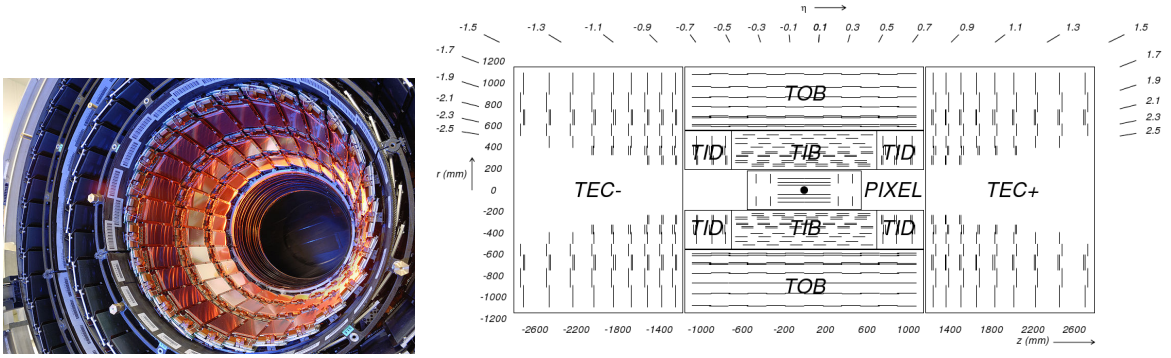


Figure 5.5: Left: Picture of silicon strips in the barrel module . Right: Schematic cross section of the tracker. Each line represents one detector module.

5.3 Electromagnetic calorimeter

Calorimeters stop and absorb the incident particles, creating an electric signal proportional to the amount of energy lost by it. It is important they are well calibrated, so the energy measured is accurate, and well segmented, so the position of the energy deposit is as well defined as possible.

The electromagnetic calorimeter (ECAL) [64] is responsible of measuring energy and position of photons and electrons, and contributes to the identification and measurement of pions and hadronic showers.

It is an hermetic and homogeneous detector of fast response and high granularity, to obtain a fast and accurate measurement of energy and position.

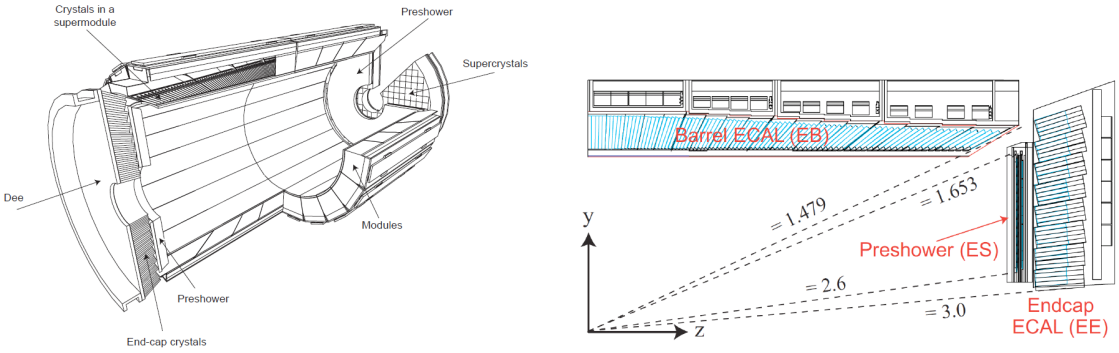


Figure 5.6: Layout of the CMS electromagnetic calorimeter. Longitudinal section of a quarter of ECAL.

It is made of 75 848 crystals of lead tugstene, $PbWO_4$, which is highly dense transparent material (8.28 g/cm^3). Electrons and photons interacting with the medium produce a short burst of light, that is collected by photomultipliers in the end of each crystal.

Radiation length of the crystals is only 0.89 cm. The Moliére radius is 2.2 cm, which is small and allows a good shower position resolution, and good shower separation due to a smaller overlap.

The response is very fast, of the order of magnitude of the LHC brunch crossing time. About 80% of the light is emitted in 25 ns. However, the amount of light emitted is low (30 γ/MeV) so photodetectors with high intrinsic gain are needed. For energies below 500 GeV, where shower leakage from the rear of the calorimeter starts to become significant, the resolution can be parametrized as

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{5.2}$$

where S= 0.028 GeV$^{1/2}$ is the stochastic term, N=0.12 GeV the noise term, and C=0.003 the constant term, according to test beam measurements.

The size of the crystals depends on the location. Only 23 cm are necessary to completely absorb the particle in the central region ($\eta = 0$) ,whereas in the endcaps they are shorter due to the existence of an additional detector, called *preshower*, which complements the detection in forward region. 61200 crystals are allocated in the barrel, grouped in 36 super modules, each of them covering half barrel length. It covers the region $|\eta| < 1.479$, and radius 1.24 < r < 1.86 m.

The granularity corresponding to a crystal front face of about 22 × 22 mm^2, equivalent to 1 degree of solid angle.

There are 7 324 crystals in each endcap, assembled in two aluminium half-disks (Dees), placed at 3.15 m from the interaction point. Each plate is divided in supercrystals, groups of 5x5 crystals. Crystals have a front face of 28.62 × 28.62 mm^2 and a length of 22 cm.

Finally, the preshower is placed in front of the ECAL endcaps with the objective of better discriminate cascades coming from individual photons, or two closely-spaced photons, which is common in π_0 decays, something very important in $H \to \gamma\gamma$ searches. Consequently, it has a much finer granularity than the ECAL with detector strips 2 mm wide.

Two planes of lead are followed by silicon strips planes, adding 20 cm thick in total. The 8 m^2 silicon strips are arranged in a grid in the endcaps to form an approximately circular shape covering most of the area of the crystal endcap.

5.4 Hadronic calorimeter

The hadronic calorimeter [65] identifies showers of hadronic particles, or *jets*, and measures its energy, or equivalently, its quadrimomenta. It requires a good energy resolution and good transverse granularity to quantify the transverse momentum. It also demands high hermeticity for a good missing energy determination.

One of the main challenges of its design comes from the size restrictions imposed by the magnet. Outermost radius in the barrel is 2.95 m, which may not be enough material to completely absorb high-energy jets, so an additional tail catcher is placed outside the solenoid to complete the calorimeter.

The detector is made of consecutive layers of brass, acting as absorber, and plastic fluorescent scintillators acting as active detector. Brass is an appropriate choice because it is a non-magnetic material. A rapid light pulse is produced by the particles, which is collected by optic fibers and led to readout hybrid photodiodes (HPDs), a type of photodetectors configured specifically to work under the magnetic conditions of the bulk of CMS. These photomultipliers create the electric signal.

In the barrel calorimeter (HB), brass planes and scintillator plastic are 50 and 4 mm thick, respectively. There are 15 planes, which corresponds to 7-10 radiation lengths. The scintillator is segmented in the η, ϕ coordinates with an amplitude of (0.087, 0.087).The HB covers up to $|\eta| < 1.3$ and is segmented in 36 wedges of 26 tonnes each, which cover 20° in the ϕ coordinate, and half of the barrel in the z-coordinate. Furthermore, the innermost and outermost plates are made of stainless steel for structural strength.

The endcaps (HE) is also divided in 36 wedges, but in this case brass plates are 79 mm thick with 9 mm gaps to accommodate the scintillators. And the segmentation is reduced to (η, ϕ)=(0.17, 0.17) from $|\eta| > 1.6$ to the maximum coverage $|\eta| > 3.0$

The outer barrel HCAL (HO), uses the coil of the solenoid as an absorber and

an additional layer of scintillator 10 mm thick to collect existing high energetic jets traversing the magnet. It is also important to identify late secondary cascades. In the central part around $\eta = 0$ a second active layer is added. Tiles in the HO are grouped in 30° ϕ sectors, in contrast to the HB with a 20° division, to match the muon chambers which are adjacent.

The forward calorimeter (HF), has to deal with a really high flux of particles. It is made of quartz fibers which detect the Cherenkov light emitted by showers. Fibers are allocated in two cylindrical steel structures with outer radius of 1.3 m and placed 11.2 m at both sides from the interaction point (IP), reaching a coverage of $3 < |\eta| < 5$. Detection of coincidences in both sides is also used for a measurement of luminosity at CMS.

5.5 Muon system

The muon system [66, 67] is allocated outside the magnet solenoid. As mentioned, a good identification and measurement of muon is vital for CMS. Hadrons and other leptons are stopped in the calorimeters, so in principle only muons cross the muon chambers, making the identification easy. Still, several sources of background have to be considered, like cosmic rays or radiation rebounded from the cavern (mainly neutrons). Only muons with $p_T > 3$ GeV reach the muon barrel instead of being trapped inside the solenoid.

The main requirements of the muon system is a high efficiency, which is achieved by a redundant system of different layers of detection, and the best possible resolution.

There are 3 different types of detectors in this system, visible in the picture of Fig. 5.7. A layout of the muon system and the location of each type of detector is depicted in Fig. 5.8.

In the barrel, drift tube chambers (DT) and resistive plate chambers (RPC) are embedded inside the iron yoke. Both are gas detectors, benefiting from the fact that the magnetic flux goes through the iron, yielding a negligible magnetic field inside the chambers. In this scenario the drift of electrons in the gas is highly linear and the position of the muon through the chamber is better determined.

The barrel system covers the range $|\eta| < 1.2$. The muon system in the endcaps covers $0.9 < |\eta| < 2.4$. The range $0.9 < |\eta| < 1.2$ is called the overlap region because both parts are present. The magnetic field is not negligible and non uniform in this area, as the field lines from inside the solenoid turn back there. A mapping of the magnetic field in the detector is shown in Fig. 5.9. Furthermore, the particle rate is significantly higher, so a different detector design is needed, and Cathode Strip Chambers (CSC) are preferred.

RPC chambers are very fast response detectors, with a resolution of 1 ns, so are employed to trigger the passage of muons. They consist of two parallel plates, a positively-charged anode and a negatively-charged cathode, both made of a very high resistivity plastic material and separated by a gas volume, working in avalanche mode

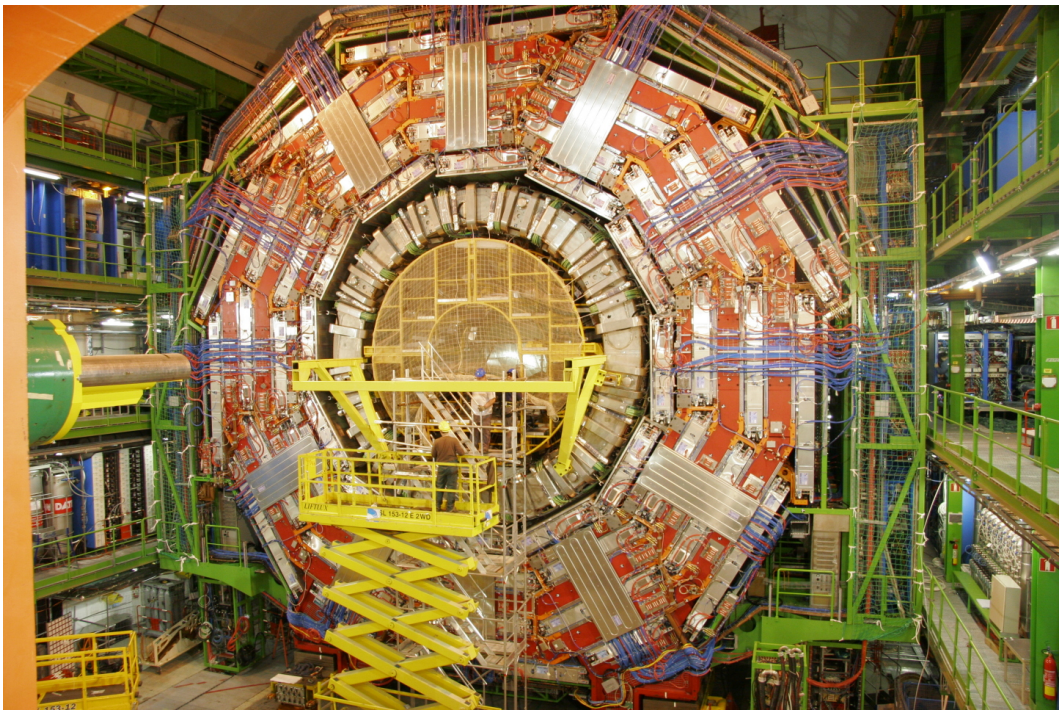


Figure 5.7: Picture of a wheel in the barrel of CMS. The steel yoke (red) and the DT chambers (silver) embedded inside are visible.

to ensure good operation at high rates. Electrons in the gas atoms are ionized by the muon and drifted to the anode, creating an electric signal. RPC are also taken to help local reconstruction on DT and CSC.

DT chambers are also gas detectors. They combine a fast response with a good spatial resolution, so they are responsible of triggering, along with the RPC, and of the spatial measurement of the track of the muon in the barrel. RPC information is also available for the spatial reconstruction, but its resolution is much worse than that obtained from the DT.

DT are composed by 8-12 layers, depending on their location, of individual cells. Each cell is filled with a gas mixture of Ar and CO_2. An electrical potential difference between an anode (in the center of the cell) and the cathode (in the edges) produces an homogeneous drift of ionization electrons followed by an avalanche close to the anode. Drift tubes description and performance will be the topic of next chapter.

The barrel iron yoke of CMS is divided in 5 wheels along the z-axis (the direction of the beam). Each one is divided in 12 sectors covering an angle of 30° in the ϕ coordinate. Each sector, in turn, has four concentric levels of chambers, called *stations*.

Six layers of RPC's are placed adjacent to the DT. Two in both sides of the two innermost DT stations, and the remaining two in the inner face of the third and fourth station.

CSC are multi-wire proportional counters (MWPC) of fast response time and finely segmented. built of arrays of positively-charged anode wires crossed with negatively-

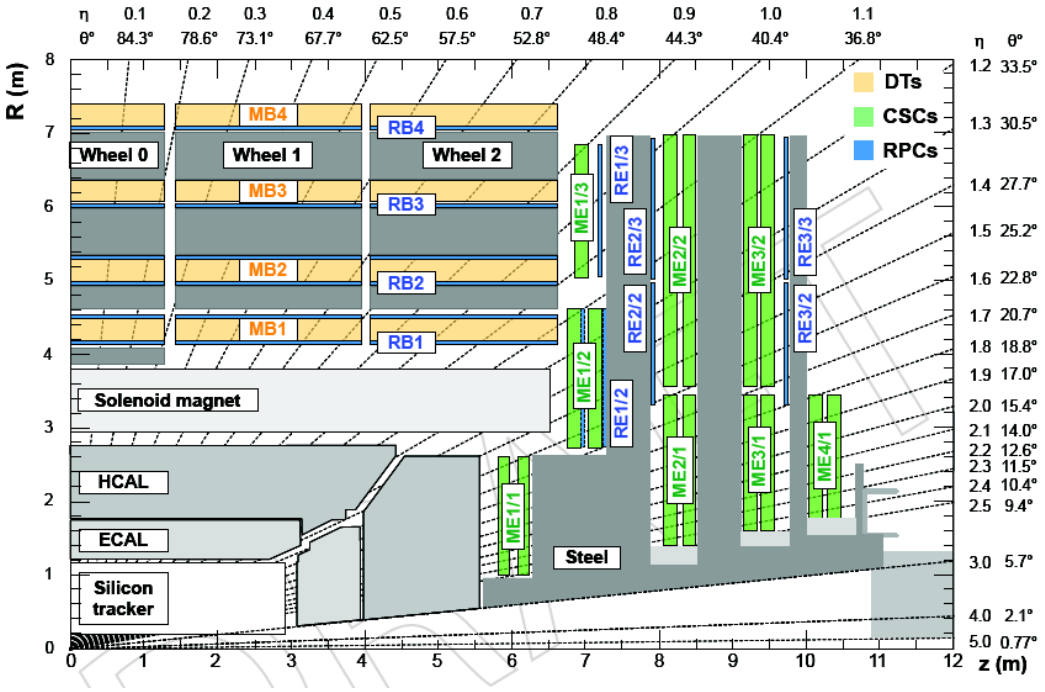


Figure 5.8: Layout of the muon system and location of the different detectors.

charged copper cathode strips within a volume of mixed gas (50% CO_2, 40% Ar, and 10% CF_4). Anode and cathode are perpendicular so a reconstructed 2-D point contains information in both θ and ϕ local coordinates. Each chamber is made of 6 gas layers, reaching a spatial resolution between 75 and 150 μm per chamber, depending on the location. It has a trapezoidal shape and covers 10° or 20° in the azimuthal ϕ coordinate, depending on its location. There are four stations or layers of CSC in each of the endcaps, each one with an adjacent RPC.

In total, 250 DT, 540 CSC and 610 RPC chambers form the muon system. Muon reconstruction is finally calculated combining the muon and tracker system, which benefits of the high resolution of the silicon detector. This imposes a very good alignment between both systems, which has been measured to be less than 160 μm.

5.6 Trigger

At the production rate of the LHC, with luminosities of near $10^{34}i$ cm^2·s^{-1} and collisions every 50 ns (to be increased in the future run of the LHC), it is not feasible the storage and the handling of all events.

On the other hand, not all bunch crossings produce processes of interest of CMS, but low energy interactions, which CMS is not interested in studying. Moreover, some of the process CMS may want to study have such high cross section that there is no need on saving all of them, but just some fraction, as not to be statistically limited by

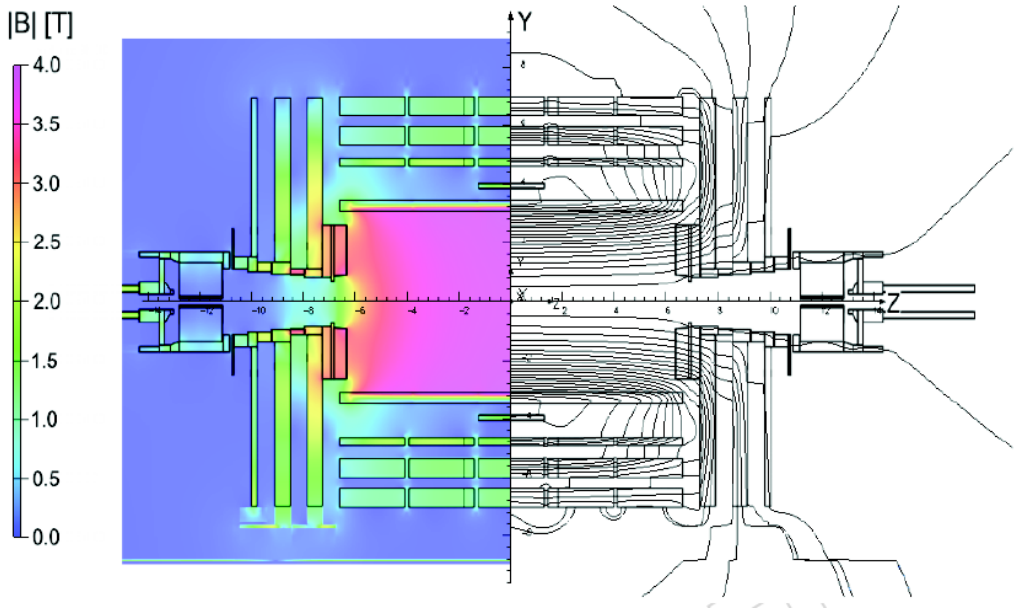


Figure 5.9: Mapping of the magnetic flux in CMS.

the size of the sample in the corresponding analysis. These factors of storage (50%, 10%, 1%...) are called *prescales*. Unfortunately, the most interesting processes (Higgs decays, New Physics ...) are the ones less frequent, ergo all effort is placed on identifying and saving all possible candidate events.

One additional difficulty is that new particles are being generated while other from previous collisions are still crossing the detector, so data is kept in pipelines that retain and process information from several interactions at the same time, and signals are assigned to a particular bunch crossing. Because of that, very good resolution and synchronization among detectors is vital.

The trigger system [68] is a combination of hardware and software algorithms to make a fast decision on whether an event is interesting and must be stored or not. The trigger privileges a partial reconstruction rather that reconstructing all possible objects, as the speed of decision is the key factor not to overexceed the buffer space.

The trigger system is divided in several steps.

The Level-1 trigger (L1) [69] (outlined in Fig. 5.10) is built of custom programmable hardware processors placed in the detectors. It is designed to reduce the initial rate of ∼400 MHz down to 50-100 kHz. Data is provisionally stored in a buffer for 3.2 μs which is the time to make a decision. In that time, the L1 uses the calorimetric and muon information. Tracker algorithm is slower so it is not used in this step. The calorimetric trigger uses the information of the ECAL and HCAL to reconstruct the best four candidates (according to their higher E_T or p_T) of the following groups: electrons and photons, central jets, forward jets and τ-jets, and the MET. At the same time, the muon trigger combines the information of all the chambers, to build the best four muons (according to their higher p_T).

If the objects satisfy certain E_T,p_T thresholds, the L1 is accepted. There are different algorithms with different thresholds, called *paths*, which may have different prescales.

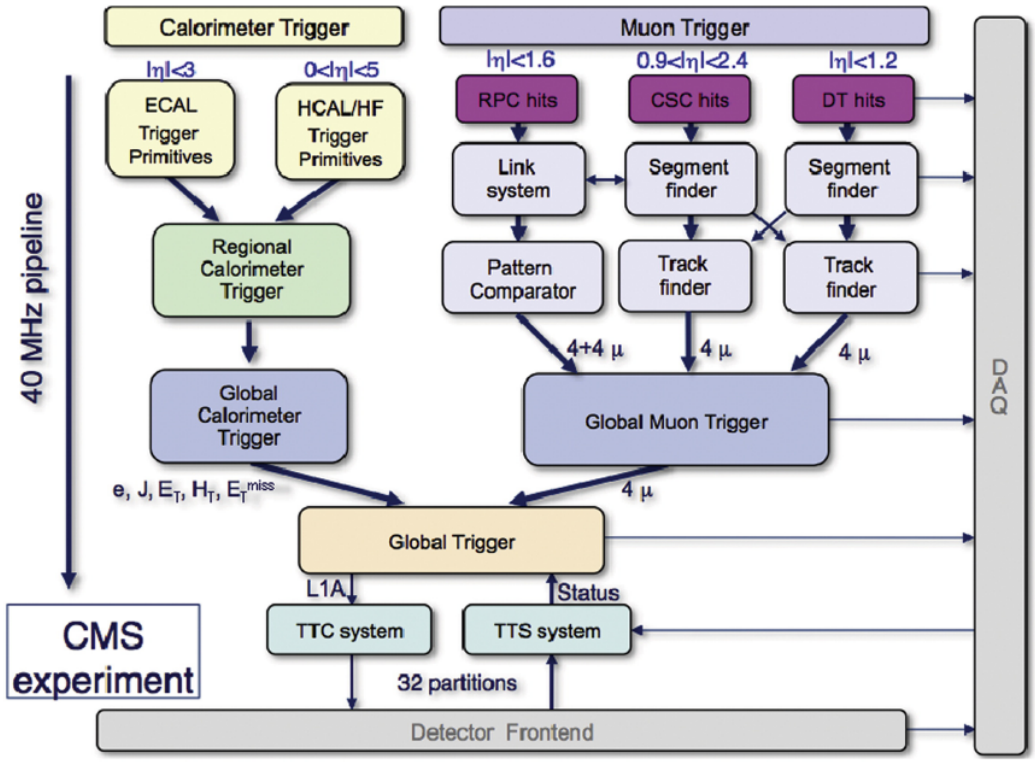


Figure 5.10: Scheme of the L1 trigger architecture.

After data-compression and zero-suppression, each event has a size of about 1.5 MB.

In case of green light from the L1, the same information is sent to the *High Level Trigger* (HLT), which is a highly flexible software system implemented in a computer farm made of about one thousand commercial processors. Regional and fast reconstruction is preferred instead of a whole reconstruction, as the only purpose is a fast (and correct) event veto. If any condition is not satisfied, the event is immediately vetoed. A set of algorithms is design to select specific event topologies, with different thresholds, an the final 100 Hz rate is achieved.

In the analysis presented in chapters 8 and 9, electron and muon triggers have been used, so they will be summarized in the following lines.

5.6.1 Electron trigger

The information of the ECAL crystals is grouped into trigger towers. A tower in the barrel is formed by a fixed array of 5×5 crystals, whereas in the endcap requires a more complex geometry. The L1 triggers the electron candidate flag in case the cascade

is contained in a narrow array of less than 2×5 crystals in the (η, ϕ) coordinates. It must be also satisfied the condition that the energy deposit in the corresponding HCAL tower does not exceed a given fraction of the ECAL energy, in order to discriminate hadronic cascades. The L1 trigger also includes the adjacent towers in the algorithm to improve the efficiency.

The HLT includes the information of the silicon detector, performing a similar, but faster reconstruction of the electron track. Pairs of pixel hits and innermost silicon strip layers are used as seeds to implement a Kalman Filter algorithm. A matching with L1 electromagnetic-calorimeter clusters is required.

Once the electron has been identified and reconstructed, several thresholds on the transverse momentum, and in the number of electrons identified, are established in different *HLT paths*.

5.6.2 Muon trigger

All three detectors, CSC, RPC and DT are involved in the muon trigger. In the L1, the CSC and DT make a local linear reconstruction in each chamber and provide position, direction, bunch crossing and quality of one muon. Quality refers to the amount of hits used in the chamber for the local reconstruction, as there may be missing hits in some layers. Then, this primitive segments are collected by the Track Finder (TF) and used as seeds to to build a standalone track with the matching segments. This step is commonly known as *L2 trigger level*. At least two stations are required in the reconstruction. In the overlap region barrel-endcap both detectors exchange information for the reconstruction.

Finally, the RPC information is included, and different p_T thresholds are required to the track for different paths.

In the L3 (HLT), the information of the silicon tracker is included. The L2 standalone is used as seed to look for matching hits in the tracker, and only a local reconstruction is performed in a slice of the tracker in a cone around the L2 track. This optimizes the time consuming in the processing with respect to a full tracker reconstruction. If the full reconstruction is achieved, the p_T resolution is significantly improved compared to the standalone track. So final requirements on the p_T of the muon are applied on the track before being accepted the the trigger.

5.7 CMS Software and Computing

After the event has been selected by the HLT farm, and stored, a full reconstruction of the event is performed. The CMS software (CMSSW) contains the code for the processing of all CMS detector output. It includes evaluation, calibration, storage, local and global reconstruction of the event, as well as all the tools used in the analysis. The ROOT Framework [70] is a software tool for statistical analysis and C++-oriented, developed by CERN. CMSSW is built on a ROOT framework. The CMSSW is created

from CMS Framework and Event Data Model (EDM), where the Event class is the key element. An Event object contains information relative to a single triggered bunch crossing, including reconstructed objects, information from the detector, etc.
CMS uses several event formats with different levels of detail and precision. RAW data includes output of the HLT. It contains information from the detector, and outputs from L1 and HLT. RECO data includes reconstructed objects (muons, electrons, jets, photons, etc. missing energy, etc.) and reconstructed hits. Analysis Object Data (AOD) is a compact format that reduces the size of the event, making large samples manageable by computing centers outside CERN.

The Worldwide LHC Computing Grid (WLCG), a distributed computing and data storage infrastructure, is needed in order to handle the enormous amount of data produced in CMS, store it and make it accessible to analyzers.

A very important tool in high energy Physics is the use of Monte-Carlo simulations (MC) to compare measured and the expected outputs. The GEANT 4 package [71] simulates the behaviour of the CMS and the interaction particle-detector. The specific details of the MC simulations used in the analysis presented here will be described in future chapters.

Chapter 6

Performance of Drift Tube Chambers in the Muon System

The detection and high precision measurement of muons is capital for CMS. Many processes of interest involve muons in their final states, like the Higgs decay channel $H \to ZZ \to \mu^+\mu^- q\bar{q}$. In fact, CMS is specifically designed for the optimization of muon detection, as its name clearly states.
In general, muons will not be absorbed by the calorimeters, as it happens with electrons, so a specific muon detection system is needed in order to identify and correctly measure its momentum. Cathode Strip Chambers (CSC) are used in the endcaps and Drift Tubes (DT) are used in the barrel. Resistive plate chambers (RPC) are used for triggering both in endcaps and barrel [3].

Hard processes, of interest for CMS, will produce central objects with high p_T, more likely to interact with barrel detectors than endcap ones. Moreover, barrel will show less occupancy thus the detection will be cleaner and easier. This chapter focuses on the description of the DT chambers, where CIEMAT has been responsible of design, construction and maintenance, and on the performance studies developed by me in the course of this PhD.

6.1 The Drift Tube Chambers

A Drift Tube chamber is a gaseous device used in the CMS barrel for triggering muons (together with RPC) and measuring its trajectory with high precision. CMS has 250 chambers (70 built at CIEMAT) distributed in the 5 wheels of the barrel. For each wheel, there are 4 concentric layers of chambers (or stations) called MB1, MB2, MB3 and MB4. For each station, there are 12 sectors covering the ϕ coordinate. Thus, each chamber covers a 30° angle in the azimuthal coordinate. As seen in Fig. 6.1, the most external station (MB4) has 2 chambers in the upper and lower sectors.

Muon system needs high acceptance and low background contamination. A robust and redundant design satisfies the first condition and improves the measurement despite

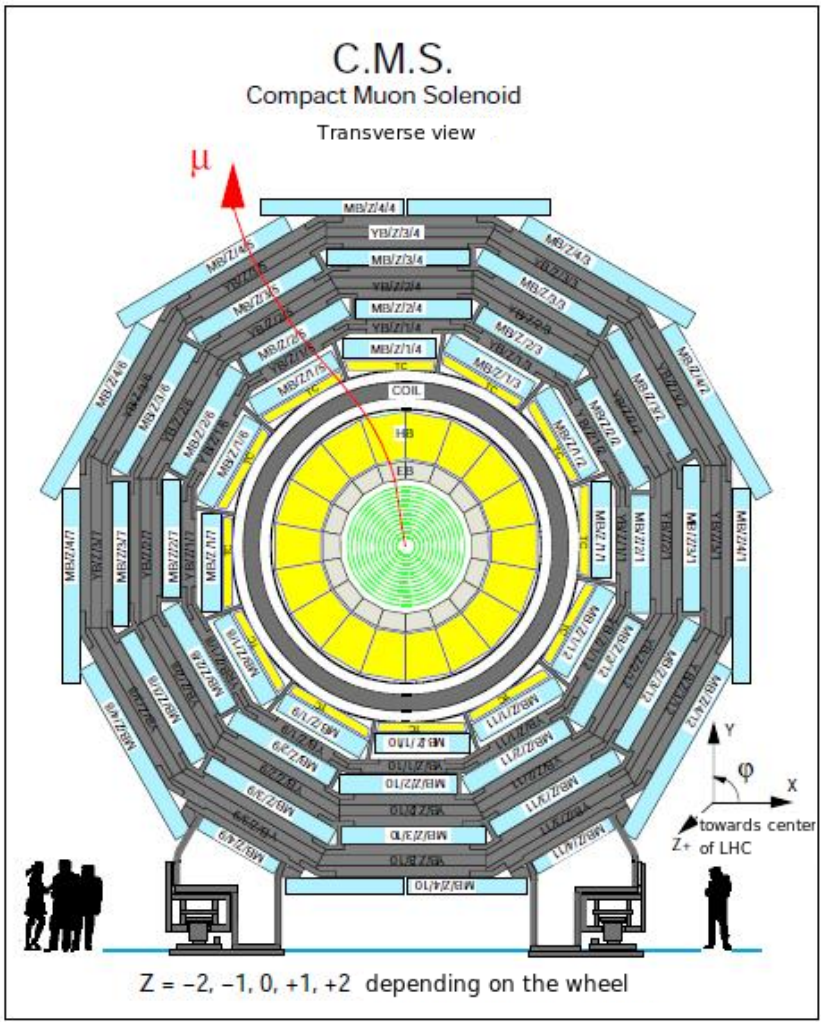


Figure 6.1: Cross-view of the CMS Barrel and the distribution of the DT chambers for one wheel.

the background level present. Chambers are located in such a way that any muon crossing one station in a non-detecting area (for example, between 2 adjacent chambers, or through the electronic readout) will necessarily cross sensitive part of the detector in the other stations.
Figure 6.2 shows the construction of the CMS detector and the installation of the DT.

6.1.1 DT Cell

The basic unity of measurement in the DT chambers is the *drift tube cell*. A cell measures single "hits", that contain information of the position of the muon track inside the cell. The drift tube cell consists on a rectangular tube of 13 mm high and 42 mm width, as shown in Fig. 6.3. In the center of the cell, a stainless steel wire of 50 μm covered of gold works as anode, for which it will be maintained at +3600 V.

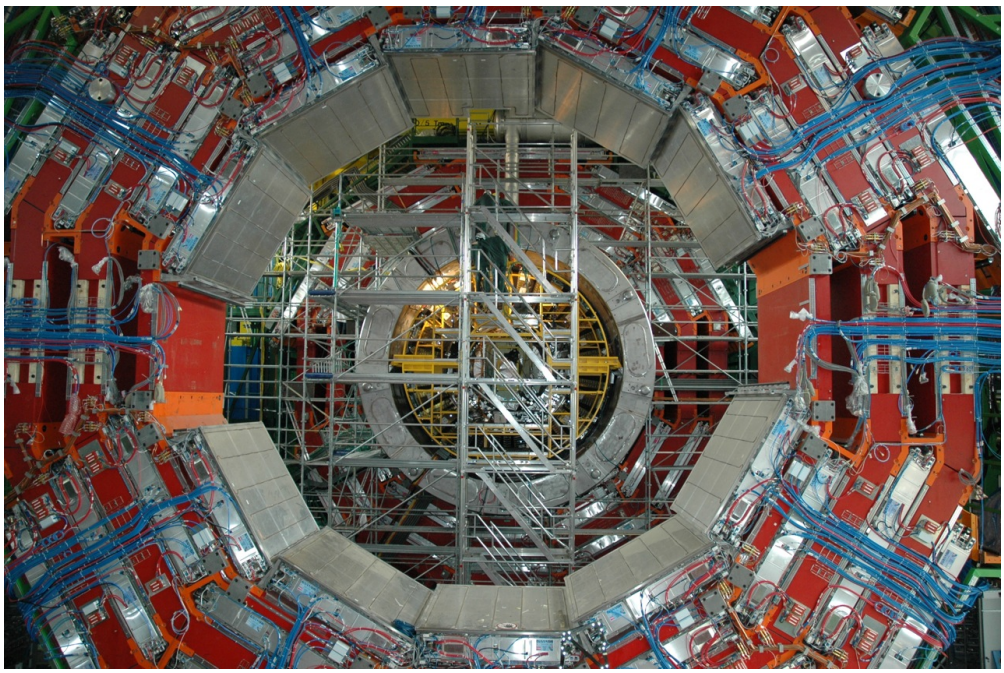


Figure 6.2: Photograph of four layers of DT (silver chambers) embedded in the steel yoke (red structure). Photograph corresponds to the construction of CMS in June 2006, when several chambers had not yet been installed.

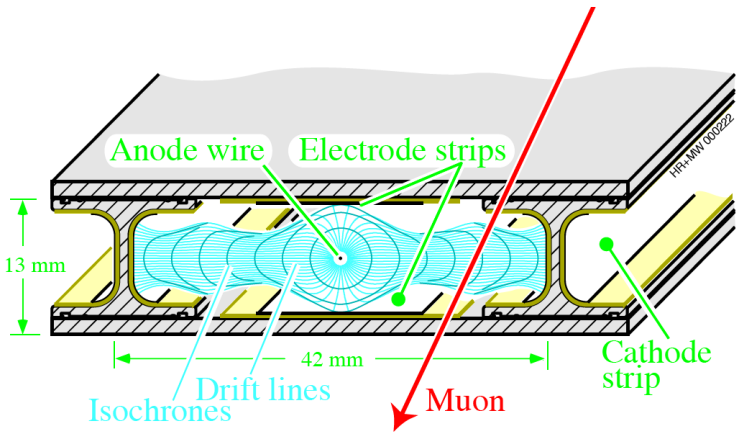


Figure 6.3: Schematic view of a DT cell.

Aluminium cathode strips are located both sides of the cell, at -1200 V. On top of that, two electrode strips at +1800 V in the upper and lower side make the drift field homogeneous. The cell is filled with a mixture gas of 85% Ar and 15% CO_2. This gas mixture succeeds in the requirements of the cell. On one hand, a high gain of the electric signal is needed to guarantee a high detection efficiency. Drift electrons only show elastic scattering in monatomic gases like argon, which results in low energy losses. This allows a high gain with a relative lower field. However, at high gain regime, excited atoms emit photons which will generate secondary ionizations in the gas, or ionize electrons from the cathode. Polyatomic gases like CO_2 have more rotational and vibrational degrees of freedom, so they efficiently absorb these photons and avoid the secondary ionizations. The mixture of the gas also determines the drift velocity of the electrons, around $v_d \sim 55$ μm/ns.

The main interaction of the muon at the energies of the LHC is the ionization of the medium. The energy losses are well described by the Bethe-Bloch formula

$$-\frac{dE}{dx} = \kappa \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2} ln\left(\frac{2m_e\beta^2\gamma^2 E_{kin}^{max}}{I^2}\right) - \beta^2 - \frac{\delta}{2}\right] \tag{6.1}$$

where $\kappa = 4N_A r_e^2 m_e$, r_e^2 and m_e are the classical radius and mass of the electron, Z and A are the atomic and mass number of the medium, E_{kin}^{max} is the highest energy transferred in each collision, I is a constant of the medium and δ is a correction term.

If the ionized electron has enough high energy, it may be ejected in another direction and produce in turn secondary ionizations in other points of the cell far from the primary ionization, called "δ-rays". This δ rays mask real hits and may degrade the reconstruction procedure, being confused with the primary ionization.

Electrons will be carried by the electric field to the anode, suffering also a diffusion process. The electric field is roughly uniform in the cell, so the drift velocity is homogeneous and the distance of the muon to the wire is estimated from a linear parameterization:

$$x = v_d \cdot t_d \tag{6.2}$$

where v_d is the calibrated drift velocity and t_d is the drift time of the electrons. Consequently, each cell provides a distance to the anode, but it is unable to determine if it corresponds to the left or right side of the cell.

In the proximity of the anode, the field is no longer linear, increasing as $V \sim 1/r$, so an avalanche is produced. The amplified signal is collected by the anode and travels through the wire to the front end electronics, where it is further amplified and discriminated above a threshold. The *High Performance Time-to-Digital Converter* (HPTDC) is the first component of the readout and converts an input signal from the front end electronics to a reference time, and afterwards, into a distance from the wire using the expression 6.2.

A time window of 1250 ns, synchronous with the LHC clock, is programmed, so the measured time of the muon hits is inside this time window. However, it does not directly correspond to the drift time of the electrons in the cell, as it includes other

factors:

$$t_{TDC} = t_d + t_0^{wire} + t_{trig} + t_{TOF} + t_{prop} \tag{6.3}$$

where

- t_0^{wire} are small delays due to the cabling and the electronics for each channel, inside a chamber. For example, θ channels have longer cables.
- t_{trig} refers to the time differences from one individual chamber to other, including the trigger L1 latency.
- t_{TOF} is the time of flight of the muon from the collision point to the cell.
- t_{prop} is the propagation time of the signal along the anode.

Once calibrated, this contributions are considered as an offset in the measured t_{TDC}. Figure 6.4 shows the timebox, distribution of the t_{TDC} in a time window of 1250 ns. The offset is about 380 ns followed by the hits produced in the vicinity of the anode. The peak is due to the strong and non-homogeneous electric field around it, producing a shift in the measured times. The peak is followed by the central region corresponding to the in-time hits from collisions, from the ones next to the anode to the ones next to the cathode. The flatness of the plateau is consequence of the linearity of the drift velocity.

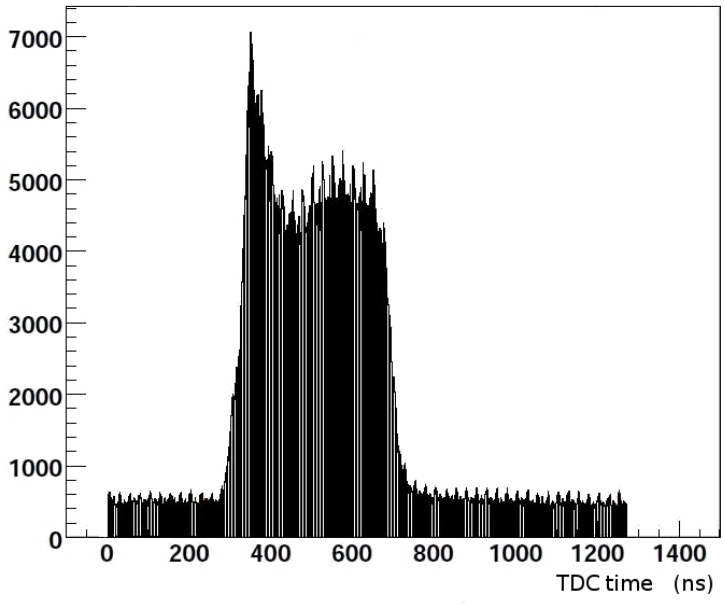


Figure 6.4: The Timebox shows the distribution of the measured time by all the hits recorded by the TDC, in a time window of 1250 ns. The first flat part corresponds to the out-of-time and noise hits. The central part corresponds to the in-time hits from collisions.

6.1.2 DT Superlayer

Cells are placed in planes called *layers*. Four parallel planes form a *superlayer* (SL). Adjacent layers are displaced half cell, to resolve the left-right ambiguity present at the cell level. On top of that, muons crossing a layer in between two cells will traverse the detection volume in the next layer. Superlayers are oriented with their wires, either parallel to the beam axis, providing information in the r-ϕ plane which allows to measure the transverse momentum p_T, or perpendicular to the axis, providing information in the r-θ plane, in the detector spherical coordinates.

A second level of local reconstruction is done fitting the reconstructed "hits" in a SL to a straight line. The result is called *2DSegment* (Fig.6.5).

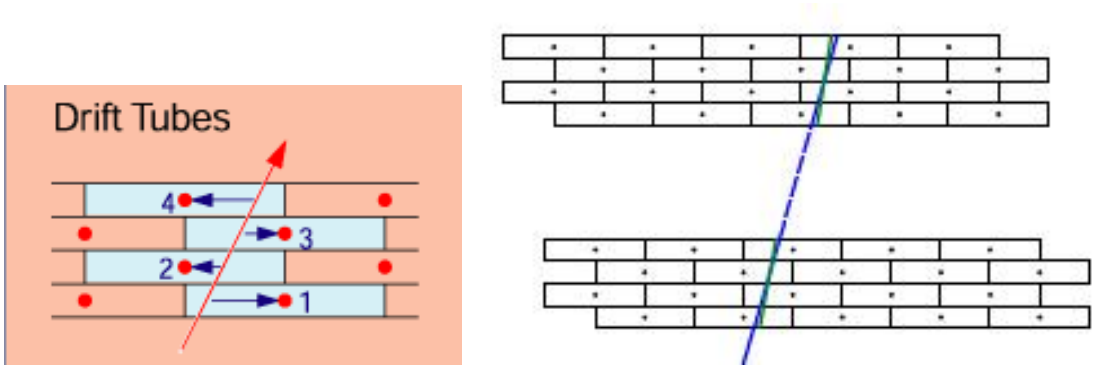


Figure 6.5: Left: Reconstruction of a 2D segment in a DT superlayer. Right: Reconstruction of a 2DSuperPhi segment in a DT chamber from the 2 superlayers-ϕ.

6.1.3 DT Chamber

A full chamber, whose layout is shown in Fig. 6.6, is made of three independent superlayers. The outermost superlayers (SL1 and SL3) measure the r-ϕ coordinate. The inner SL measure the orthogonal r-θ coordinate, although the outermost stations (MB4) lack of the θ-superlayer. A honeycomb structure of 22 cm is placed between the first and the second SL to provide stability and rigidity to the chamber, and to increase the level arm between the two ϕ-SL in order to improve the measurement. In addition, it shelters the so-called *minicrates*, where part of the electronics is placed. Chambers are limited in size in the longitudinal dimension by the segmentation of the barrel yoke, and are about 2.5 m long. On the transverse side, their length varies with the station, ranging from 1.9 m for MB1 to 4.1 m for the largest MB4.

The third step in the local reconstruction is to match and refit the two ϕ 2DSegment (if both exist) into one 2DSuperPhiSegment, as the one illustrated in the right plot of Fig. 6.5. Finally, the θ-SL information is added to obtain the lineal 3D linear segment with information of position and direction in the (ϕ,θ) coordinates, called *4DSegment*.

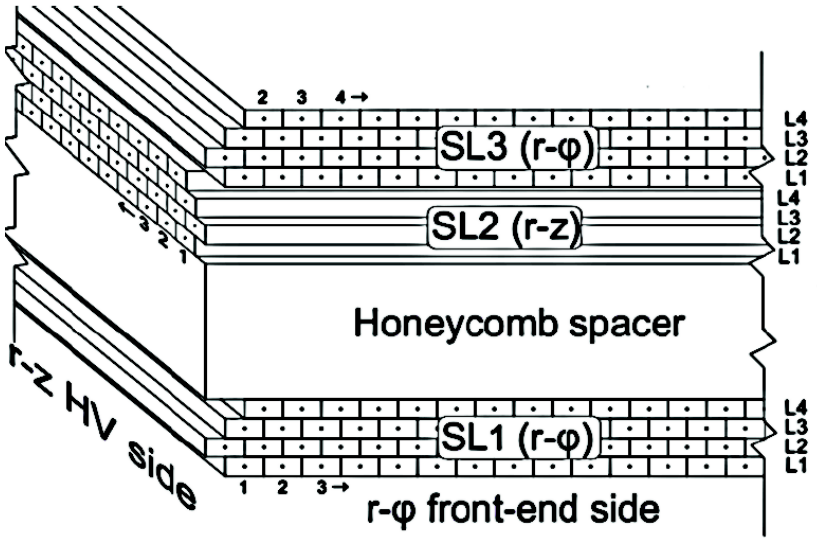


Figure 6.6: Layout of a Drift Tube Chamber.

6.2 Performance of the Drift Tubes

With the first data from collision at $\sqrt{s}$ =7 TeV recorded by CMS, a strong effort has been done to study the performance of the detector, and in particular the DT's. An accurate understanding of the response of the detector with real data is key for a good reconstruction and interpretation of the data.

Several studies are presented below concerning the efficiency, resolution and noise in the DT cells and chambers. The instantaneous luminosity in the LHC has gradually increased with time. An integrated luminosity of $\mathcal{L}$=2.6 pb^{-1} recorded by the first half of 2010 has been used for the resolutions. The full 2010 dataset (40 pb^{-1}) has been used for the efficiency and noise measurements.

Due to the low luminosity existent by the time, the cosmic muon background is also analysed, as crosscheck.

An inclusive sample with very loose selection requirements (*"Minimum bias"*) has been used, selecting all reconstructed 4D segments, whether they belong or not to an afterwards reconstructed muon.

6.2.1 Efficiency

Efficiency of measuring a single hit has been measured. The efficiency of a chamber as a whole to reconstruct a segment has not been considered, but the efficiency of a particular layer or cell. The sample consists on the reconstructed 4D segments, in order to map the hits associated to it and look for inefficiencies. An "ideal" segment should contain 12 hits (1 per layer), or 8 hits if it belongs to the MB4 station (which, as mentioned before, lacks of $\theta-$SL). However, layers are sometimes inefficient or contain

more than one hit used in the reconstruction. Figure 6.7 shows the distribution of number of hits associated to a segment. Some segments contain even more than 12 hits. It may happen for tracks with high impact angle, which originate hits in adjacent cells of the same layer. The distribution pattern is due to the reconstruction algorithm. A 2D segment is built from at least 3 out of 4 layers in a SL. Consequently, the peaks at 4, 8 and 12 corresponds to segments reconstructed from 1,2 or 3 2D segments, respectively.

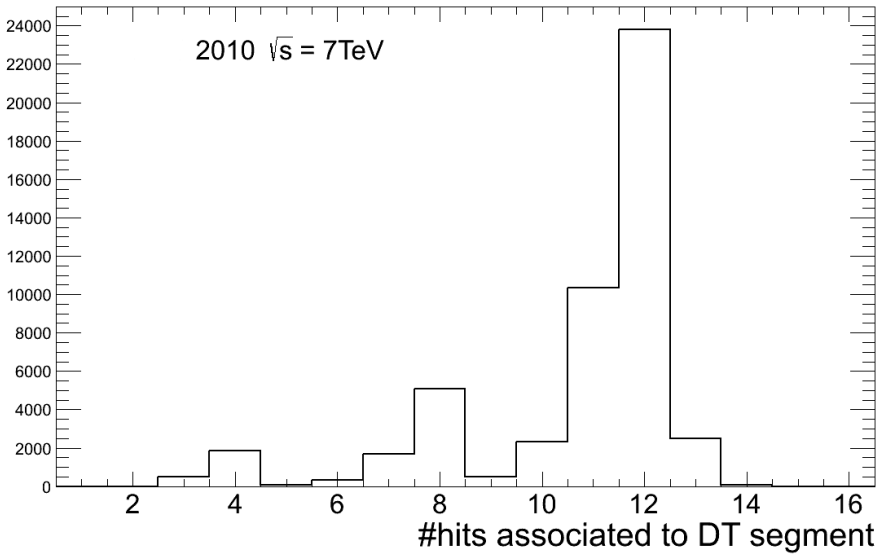


Figure 6.7: Distribution of number of reconstructed hits in a DT 4D segment.

Certain quality cuts are necessary not to bias the calculation because of background contamination. For example, low energy cosmic muons leave signal in the outer layers of the MB4 station, but are usually absorbed in the yoke. Also, high energetic hadronic showers may not be totally absorbed in the calorimeter and reach the innermost layers of the most internal station. This is called "*punch-through*". On top of that, and for geometrical reasons, segments with a big impact angle may traverse a chamber only partially, leaving it by one side. All these effects must be taken into account.

During the first semester of 2010, luminosity was low and only 2.6 pb^{-1} were recorded. Therefore, a very inclusive sample was used in order to manage reasonable statistics in my master thesis project [2], but harder quality cuts were needed to remove the background. The measured efficiency was dependent on the selection criteria used. Figure 6.8 depicts the average efficiency per layer considering different quality cuts, starting from no cuts (first bin) to demand at least 7 hits in the ϕ-SL, 4 hits in the θ-SL and a ϕ <25°impact angle (last bin).

By the end of 2010, the recorded luminosity reached to 40 pb^{-1} and an enriched sample of W and Z decays was available for analysis in the final published update [3], so quality cuts were also adjusted. In the following efficiency results, performed with this muon enriched sample, at least 5 (3) hits were required, located in at least 4 out

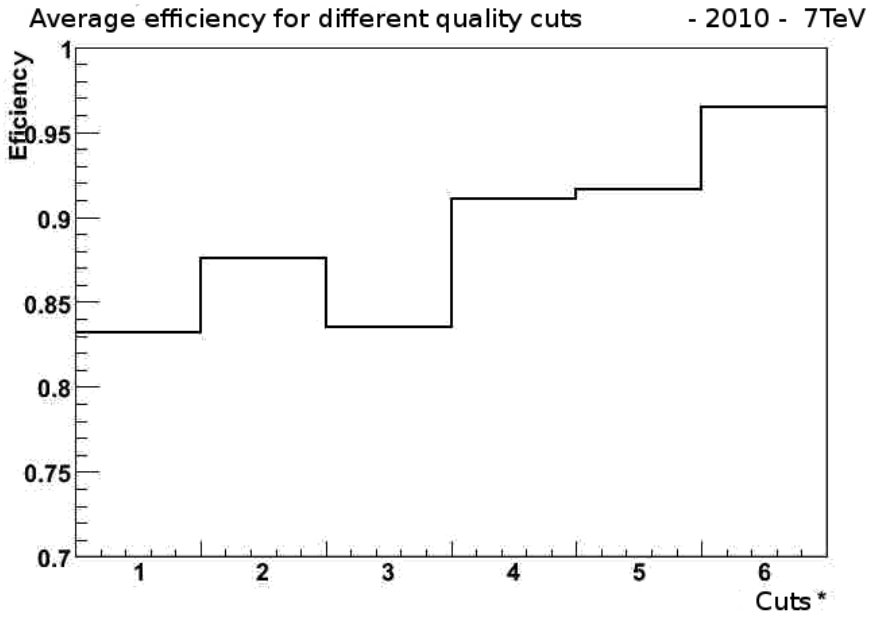


Figure 6.8: Average efficiency considering different quality cuts: Bin1: no cuts. Bin 2: require any hit in at least 2SL. Bin 3: require ϕ and θ components in the segment. Bin 4: require any signal in at least the two ϕ SL. Bin 5: require at least 7(out of 8) hits in the ϕ-SL and a χ^2 <20 in the fit. Bin 6: require at least 7 hits in the ϕ-SL, 4 hits in the θ-SL and a ϕ <25 degree impact angle.

of 8 (3 out of 4) chamber layers, corresponding to the r-ϕ (r-θ) plane. On top of that, a local inclination angle of ϕ <40° in the transverse plane to the beam axis was also required.

Using the set of hits associated with a reconstructed segment, the segment was fitted again, once per layer, ignoring the information for that layer. Therefore, the position of the segment in the layer under study is determined in an unbiased way. A hit is looked right after in that position.

The present study tried to measure the efficiency of the detector device, besides the reconstruction fit. Because of that, two kinds of efficiencies are considered. The "single hit" efficiency is the efficiency to find a reconstructed hit within a cell , whereas "associated hits" efficiency refers to the efficiency to actually associate a hit to the segment. The latter efficiency is by definition lower, as it includes the effects of the calibration and fitting procedures. Noise hits, $\delta-$rays or bad reconstructed hits providing a large χ^2 fit may be removed from the segment.

On the other hand, for the "single hit" efficiency, on top of the existing "associated hits", the segment is extrapolated to the layers without hit and individual single hits are searched around that position, without further constrain.

A per layer efficiency has been computed and shown in Fig. 6.9. The plot on the top compares the efficiency for "single hits" and "associated hits", being the former 97.7% on average, and the latter of 96.1%, a little bit lower as expected. Plot on the bottom

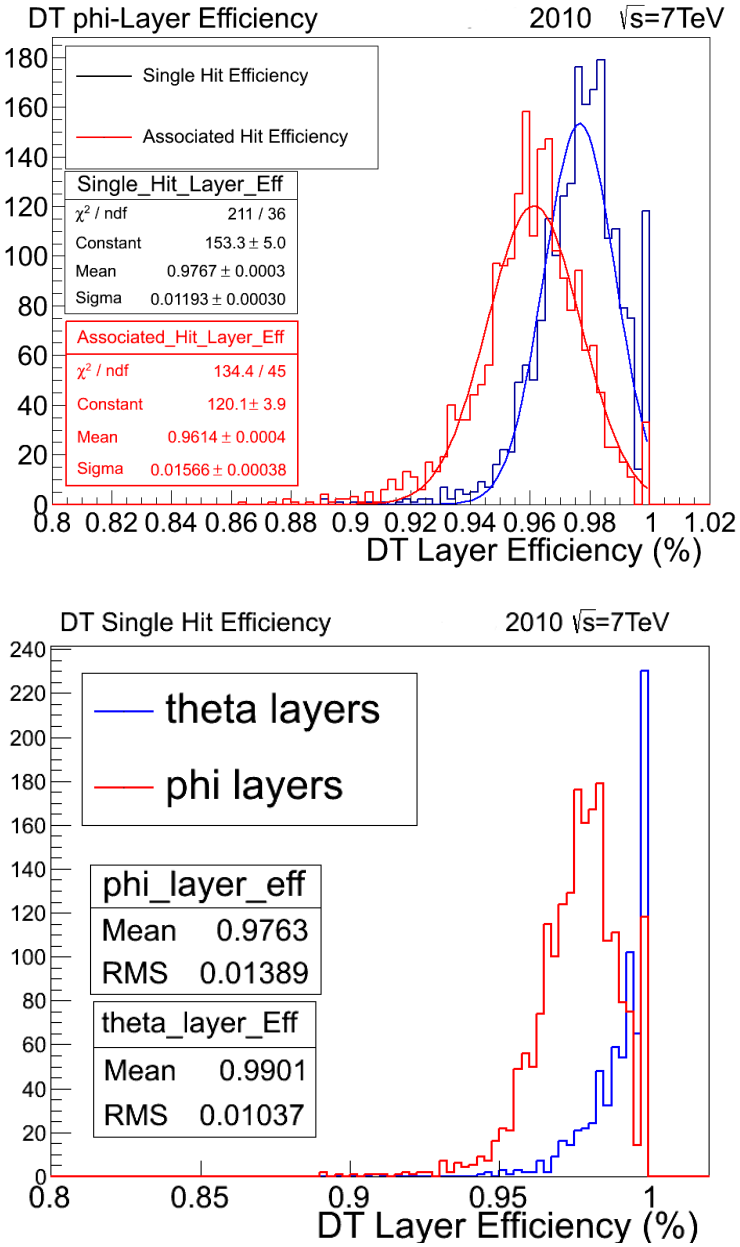


Figure 6.9: Top: Distribution of DT layer efficiency fitted to a Gaussian, for "single hits" (black) and "associated hits" (red). Bottom: Distribution of DT layer "single hits" efficiency splitted in r-ϕ (red) layers and r-θ-layers (blue).

in Fig. 6.9 compares the ϕ and θ layers. The difference between them correspond to the tighter requirements (3 hits out of 4) to reconstruct the θ component. In both plots, the entries in the bin $e_{ff} = 1$ correspond to layers with very low occupancy.

Different levels of granularity are considered. Fig. 6.10 maps the efficiency per superlayer for all SL in the detector. An overall homogeneous behaviour is observed through all the muon barrel detector.

It is also possible to study the sources of inefficiency. Plot in the top in Fig. 6.11 shows the hit reconstruction and the hit association efficiencies as functions of the position in the layer for a subset of DT MB1 chambers. The efficiency is approximately constant along the layer and the cell structure is clearly visible. The main factor of inefficiency is geometrical, when a muon crosses through the cathode I-beam that separates two adjacent cells. However, this does not imply a loss of information due to the redundancy of the muon reconstruction.

Plot in the bottom in Fig. 6.11 shows the efficiency as a function of the position inside a DT cell. Apart of the drop in the edges of the cell due to the I-beam wall, the efficiency is $\geq$99% in all the detection area. The association hit efficiency is up to 2% lower mainly due to the contribution from δ-rays. δ-rays may create an early hit that masks a good one because of the electronics dead time. In fact, the hit association efficiency matches the reconstruction efficiency in the central region of the cell, close to the wire, where this effect due to δ-rays is smaller.

6.2.2 Resolution

Hit position resolution is determined from the 4Dsegments, from an inclusive sample with very loose requirements (*"Minimum bias"*), selecting all reconstructed 4D segments, whether they belong or not to an afterwards reconstructed muon. The muon enriched sample had too low statistics for the considered luminosity. One advantage of using local segments is that the relative alignment of chambers does not affect the result.

Good quality segments are required, to guarantee the use of well reconstructed segments and avoid any bias from possible errors in the muon reconstruction process. Only segments with 7 (4) hits in ϕ (θ)-SL and local angle ϕ <25° are selected. Two different methods are studied and crosschecked:

The meantimer method

The meantimer [72] is a magnitude that measures the maximum drift time T_{max}, from the cathode and the wire. It makes use of 3 out of 4 of the individual times measured in a SL. Depending on which 3 layers are used, two possible definitions of meantimer are possible:

$$MT_1 = \frac{1}{2}(t_1 + t_3) + t_2 \tag{6.4}$$

$$MT_2 = \frac{1}{2}(t_2 + t_4) + t_3 \tag{6.5}$$

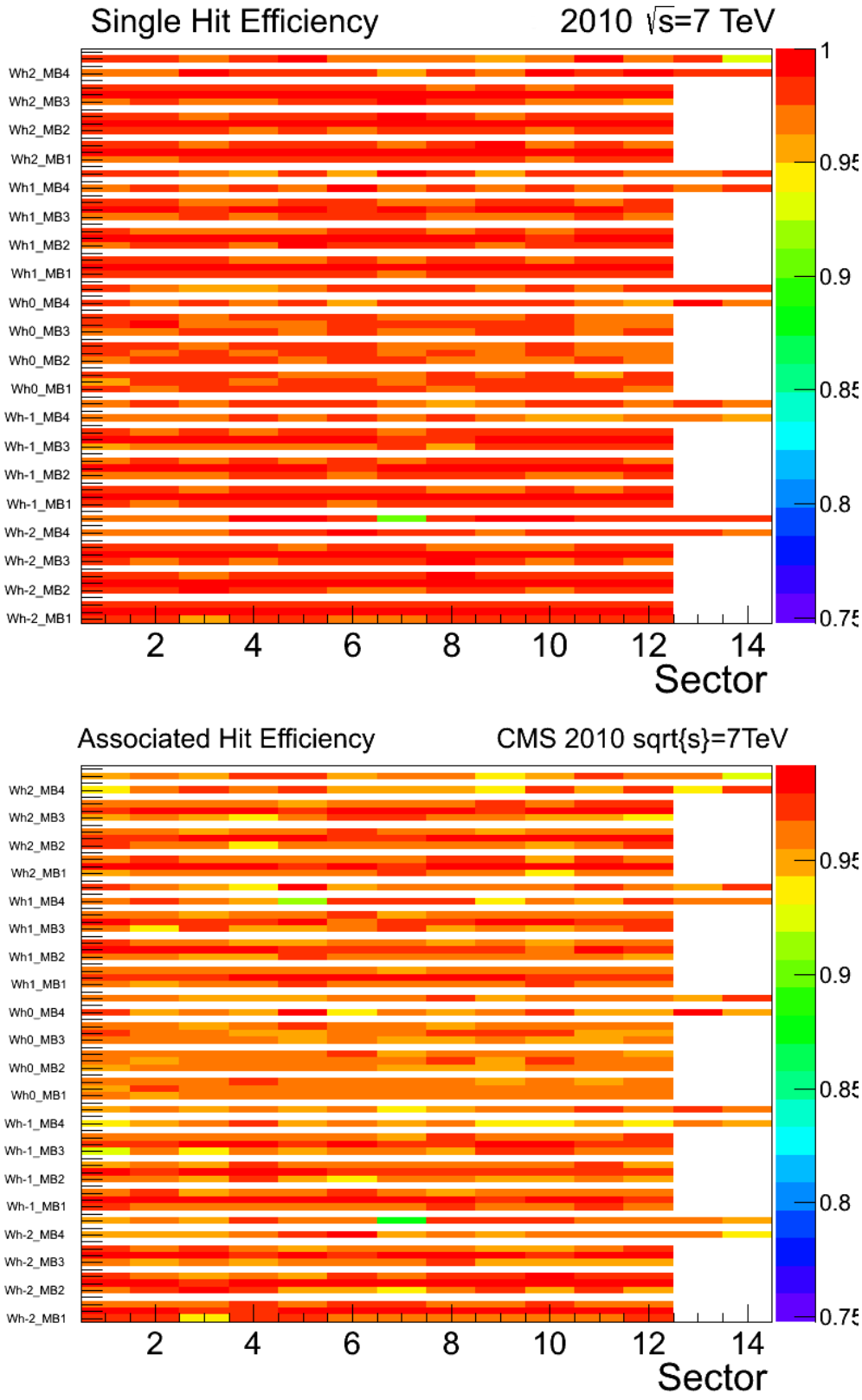


Figure 6.10: DT superlayer efficiency map for "single hits" (top) and "associated hits" (bottom). X-axis represents the sector, and the Y-axis accounts for the wheel (wh) and station (st).

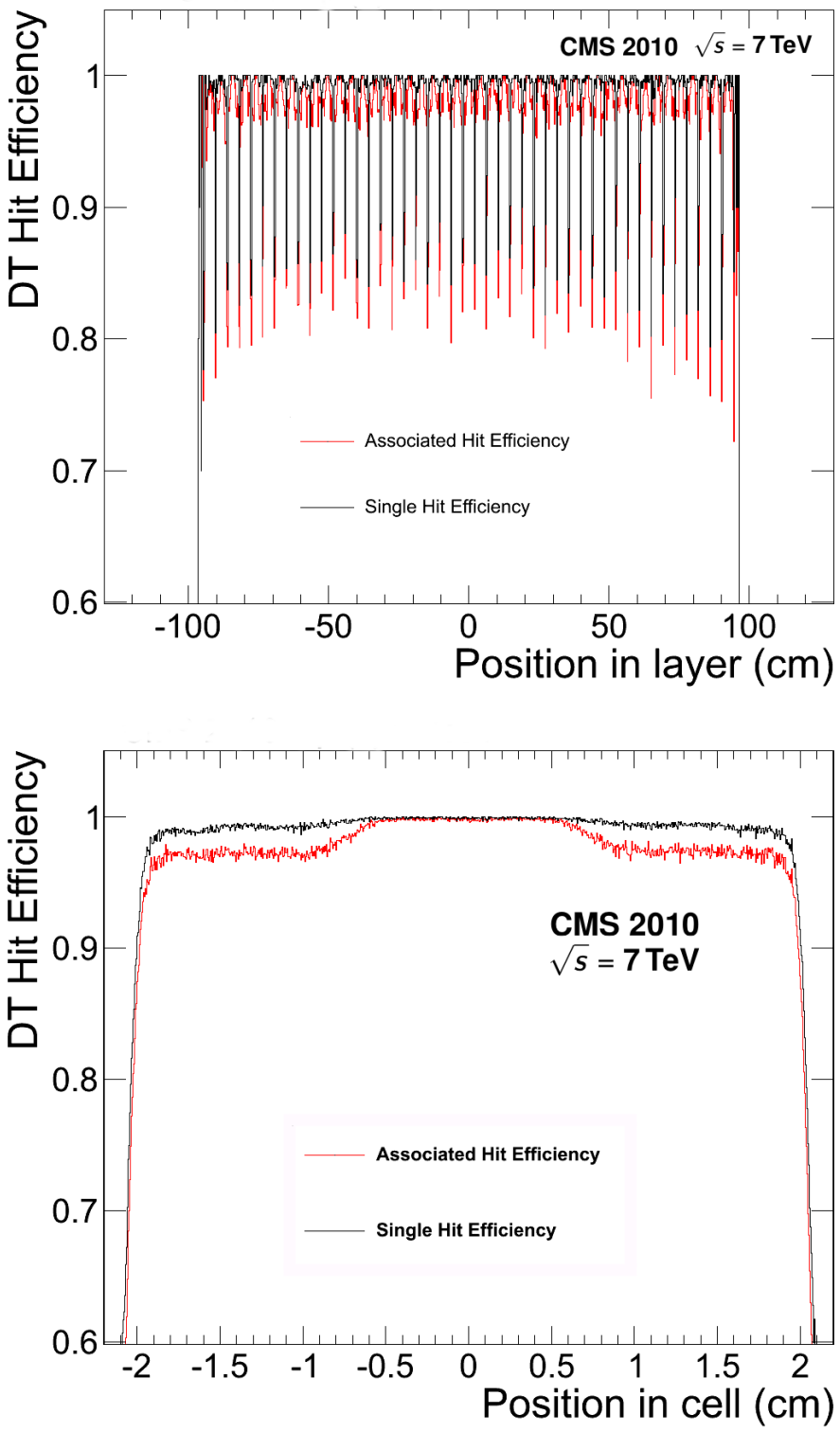


Figure 6.11: Top: Single hit reconstruction (black) and association (red) efficiencies as a function of the track position in a DT MB1 layer. Bottom: Single hit reconstruction (black) and association (red) efficiencies as a function of the track position in a DT cell.

where $i = 1...4$ is the number of layer.

Assuming the resolution is the same in all 4 layers, it is related to the dispersion of the meantimer via the expression:

$$\sigma_y =< v_d > \sqrt{\frac{2}{3}}\sigma_{MT} \tag{6.6}$$

where v_d is the drift velocity of the electrons in the cell, which is also estimated from the meantimer value:

$$v_d = \frac{L_{cell}}{2 < T_{max} >} \tag{6.7}$$

The meantimer is only valid when the muon crosses the cells by different sides of the anode in consecutive layers, which occurs when the local angle is not too high. This is the reason of the cut in ϕ. Otherwise the sum of times is not equivalent to T_{max}. Figure 6.12 shows different examples of the meantimer for different scenarios.

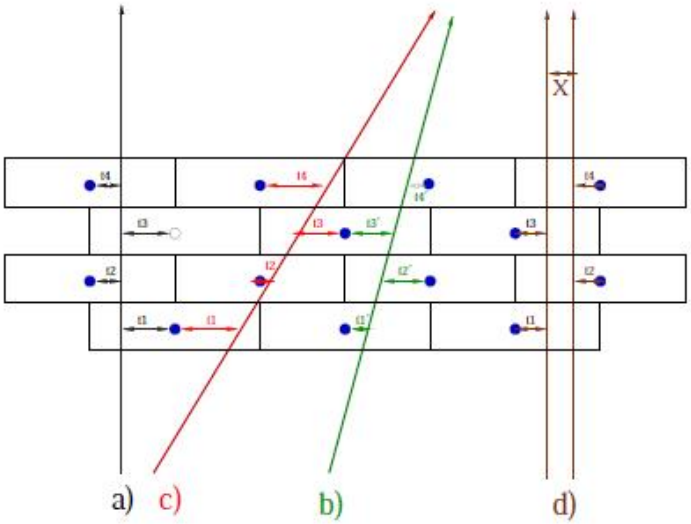


Figure 6.12: Different muon scenarios for the DT meantimer determination: muon track perpendicular to the chamber (a), muon track with a certain angle (b), muon track with high angle, crossing 2 consecutive cells by the right side, and the case of two parallel muons crossing the same cell (d).

A Gaussian fit to the meantimer distribution is performed chamber by chamber. Fig. 6.13 shows the mean value of T_{max} measured for the ϕ and θ components of the 250 chambers. The distribution oscillates from 380 to 410 ns. The second peak between 400 and 410 ns corresponds to the MB1 chambers, in the external wheels. In that region the magnetic field is not negligible, affecting the drift velocity and, therefore, T_{max}. On top of that, the incident angle of the muon also affects the drift velocity. Muons cross chambers in external wheels ± 2 with a higher angle than central wheels, which affects in different way the ϕ-SL and θ-SL, as it will detailed in section 6.2.2.

Taking into account that the half size of a cell is 2.1 cm, from Fig. 6.7 an average drift velocity is estimated to be $< v >_d = 53.7 \pm 0.5$ μm/ns, consistent with previous tests performed with cosmic muons [73].

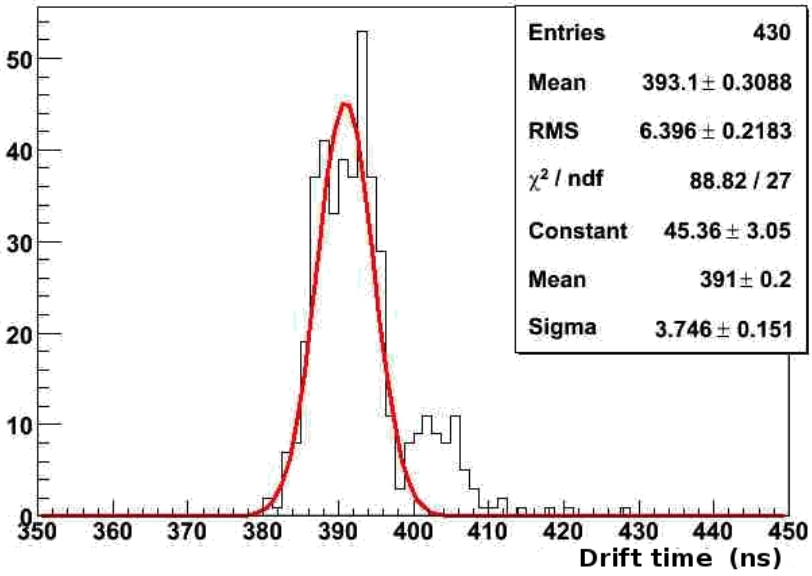


Figure 6.13: Maximum drift time in a cell, for all chambers, calculated from the meantimer method. ϕ and θ components are considered independently. The second peak between 400 and 410 ns corresponds to the MB1 chambers, in the external wheels ± 2, where the magnetic field is not negligible and the incident local angle is higher.

The time of propagation of the signal along the anode t_{prop} depends on the position where the muon crosses the chamber. Figure 6.14 shows the meantimer of ϕ SL as a function of the position in the θ-SL. A delay of $\sim$10 ns is observed from tracks placed near one border of the chamber with respect to tracks crossing in the other side, near the readout. The T_{max} is corrected with this information, except for the MB4 station where no θ information is available.

Finally, from Eq. 6.6 and from the standard deviation σ_{MT} of the meantimer fits, the value of the average resolution is obtained.

The residual method

Hit resolution is determined from the distribution of hit residuals with respect to the segment. The residual is defined as the distance of the hit position to the position of the segment at the corresponding layer

$$res = y_i - (A + B \cdot x_i) \tag{6.8}$$

The standard deviation of the residuals is linearly correlated with the resolution

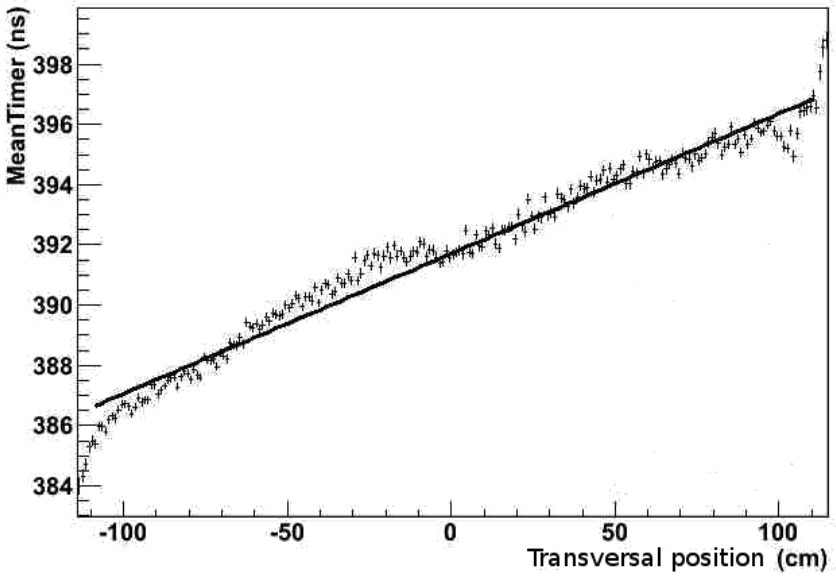


Figure 6.14: Meantimer value in the ϕ SL as a function of the position in the θ-SL. The linear slope is the result of the propagation of the signal through the wire.

[72]:

$$\sigma_y = \sqrt{\frac{N-1}{N} - \frac{x_k^2}{\sum x_i^2}} \cdot \sigma_{res} \tag{6.9}$$

where N is the number of points used in the fit (8 in the case of ϕ segment and 4 in the case of θ segments), x_k is the position of the k-th plane in a local reference frame, and σ_{res} is the dispersion of the residual.

Resolution results

The mean resolution of the position in the cell for the meantimer method is σ_{y_ϕ}=378.6±52.5 μm for the ϕ-SL and σ_{y_θ}=460.8± 71.6μm for the θ-SL. For the residual method, the resolution obtained is higher: σ_{y_ϕ}=524.2±48.6μm and σ_{y_θ}=733.7±193.9μm. This result has turn out to be worse than expected, and worse than the actual behaviour of the detector. A discussion about the causes of it will be presented at the end of this section.

First of all, the worse σ_{y_θ} compared to σ_{y_ϕ} is because there is only one θ-SL per chamber, so the statistics in much lower. Figures 6.15 and 6.16 show the ϕ resolution measured for all the chambers, for the meantimer method and residual method, respectively. Important differences have been observed among stations, which explains the pattern of the histogram, so it has been splitted into stations and fitted to Gaussian, which are shown in the figures as the coloured fits.

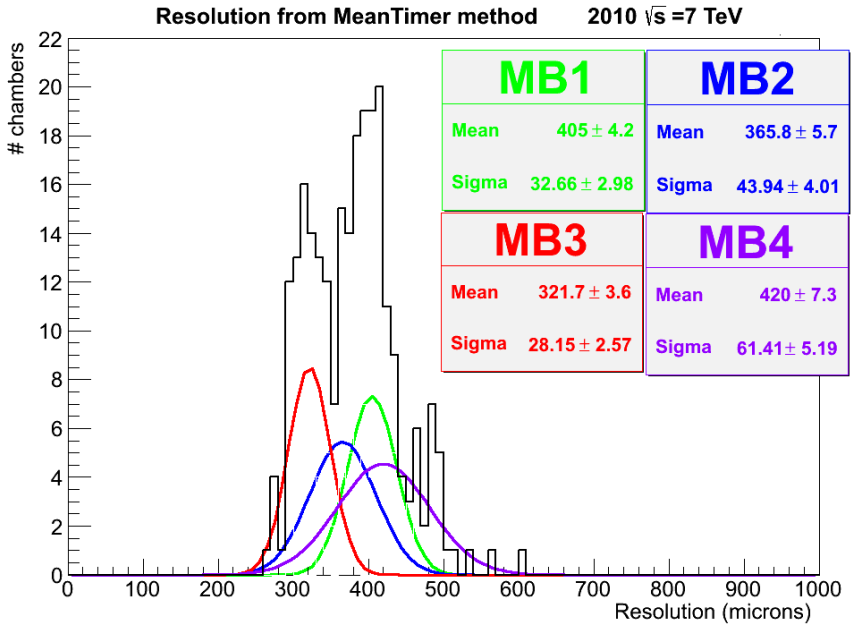


Figure 6.15: Resolution in the ϕ component from the meantimer method for all chambers (black histogram). Differences are observed as function to the distance to the interaction point, so Gaussian fits are performed to the distributions splitted in the different stations.

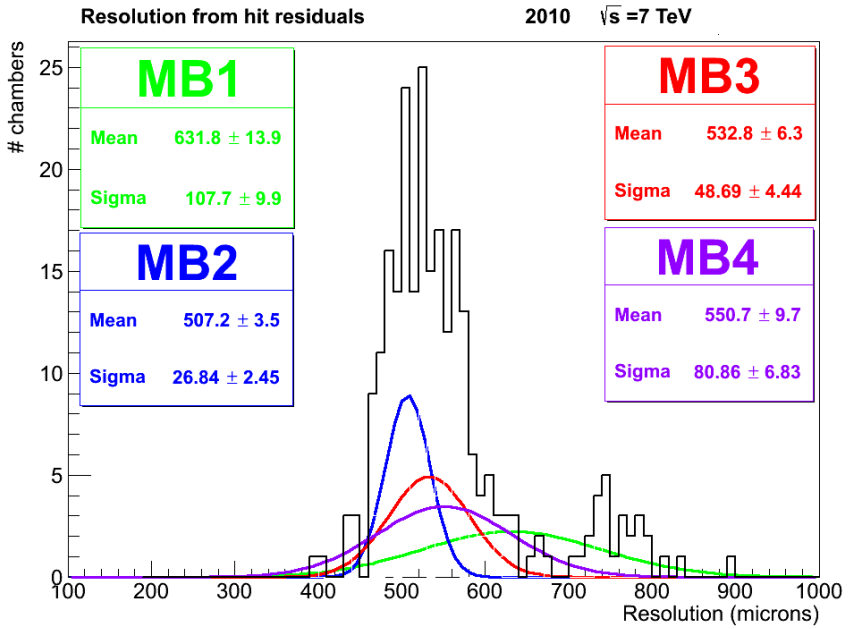


Figure 6.16: Resolution in the ϕ component from the residual method for all chambers (black histogram). Differences are observed as function to the distance to the interaction point, so Gaussian fits are performed to the distributions splitted in the different stations.

Resolution is higher in MB1 due to the non-negligible magnetic field, especially in the wheels ±2. MB4, on the other hand, lacks of enough statistics, as only higher-energy muons will reach the outermost stations. On top of that, they suffer from all the cosmic and surrounding radiation. MB2 and MB3 are the ones better protected against background, magnetic field, and with more decent statistics.

Figure 6.17 shows the average resolution obtained as a function of the radial distance (stations) and longitudinal distance (wheels) to the interaction point. Detector shows a symmetric behaviour with respect to the beam axis. Central part of the barrel has worse resolution because the statistics is lower. Only muons with high p_T reach wheel 0. Furthermore, perpendicular cosmic muons crossing wheel 0 will be misidentified as muons coming from collisions, worsening the quality of the subsample. On top of that, the θ angle affects directly the resolution. In ϕ-SL , larger angles, present in external wheels, slightly improves the resolution because they increase the path inside the cell, producing more ionizations. The opposite effect occurs in θ-SL, because in this case it is the angle in the measurement plane, so deviations in the angle produce deviations in the linearity of 6.2. In Fig. 6.18 the resolution in the θ-SL calculated from the residual method is presented as a function of the θ angle. Resolution increases with this variable because CMS drift tubes are not designed to perform high angle measurements.

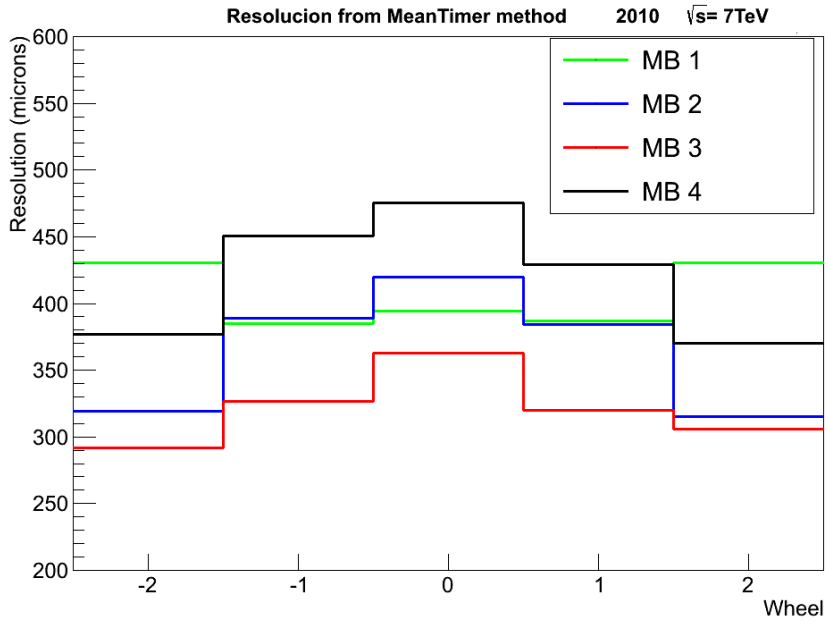


Figure 6.17: Average ϕ resolution obtained as a function of the radial distance (stations) and longitudinal distance (wheels) to the interaction point.

The residual method gives a much poor result. This is mostly due to the fact that correction for the t_{prop} of the wire has not been done, as for the meantimer, as it would require a full re-reconstruction of all segments. The variation of measured times produced by t_{prop} alters the track reconstruction and consequently the resolution.

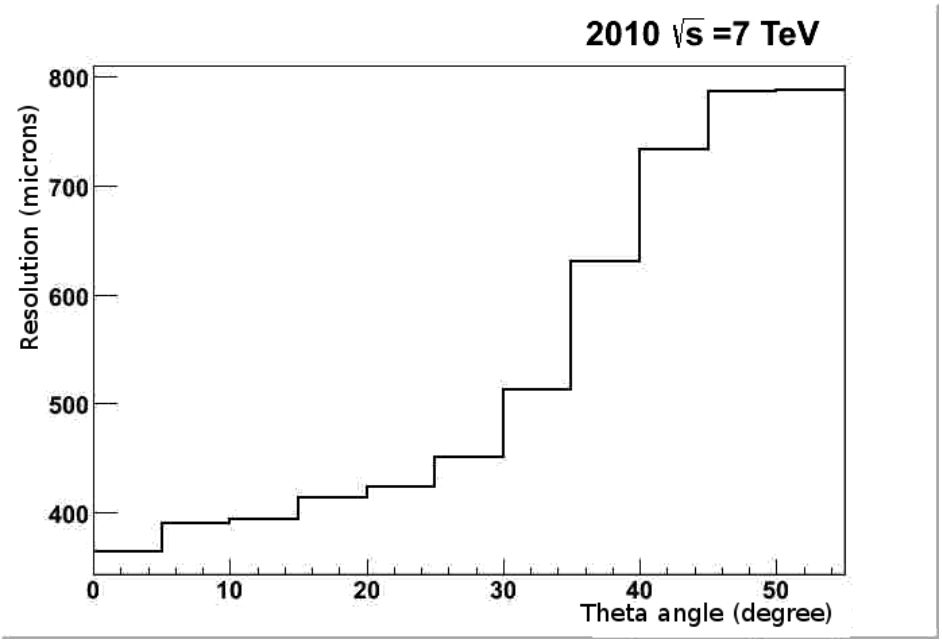


Figure 6.18: Average resolution of the θ-SL as a function of the local θ angle.

Current CMS reconstruction software automatically includes this correction, but it was not included at the beginning of 2010.

Even though, the resolution obtained is higher than expected (about 200-300 μm)[72, 73]. Several factors must be taken into account, related to the conditions of the beginning of the LHC:

- Really low statistics accumulated, with a large muon background contamination due to the low instantaneous luminosity.

- Synchronization among chambers was not perfect and was continuously changing and improving. Consequently, differences of several nanoseconds exist in data from some running periods to others, for each chamber.

- Calibration and alignment were not perfect and were changing through time.

In conclusion, a much larger sample and a reprocessing of the data with better synchronization, calibration and reconstruction was needed to improve the results. This has been done in 2011 [3] where resolution results improved to (220-350) μm in ϕ-SL and (266-785) μm in θ-SL, in agreement with the TDR expectations [66].

Finally, it must be said that the spatial resolution in the whole chamber was obviously much better than the resolution for a single hit, although was not studied in the present work.

6.2.3 Noise

As mentioned, DT chambers suffer from background radiation. Typical sources of background are muon cosmic rays, secondary muons from hadronic decays, punch-through hadrons, and photon and neutrons decays. Hits produced in previous or subsequent collisions and assigned to the wrong bunch crossing are also a source of background, as well as a source of inefficiency in the correct event they should be assigned to. It is important to understand the background radiation levels in the overall DT performance, as it directly affects the trigger performance, the pattern recognition in the muon reconstruction and the aging of the detector.

The collection of all hits recorded out of the trigger time window do not correspond to real data from collisions, except for the hits assigned to the wrong bunch crossing, and are used to analyze the background without the "true collision hits" background. In the timebox distribution, depicted again in Fig. 6.19, they correspond to the pedestal. Two different tails, marked in red, are considered independently and crosschecked, showing both the same results. The first Time Window (TW1) covers 0< t_{TDC} <250 ns and the second one (TW1) covers 830< t_{TDC} <1250 ns.

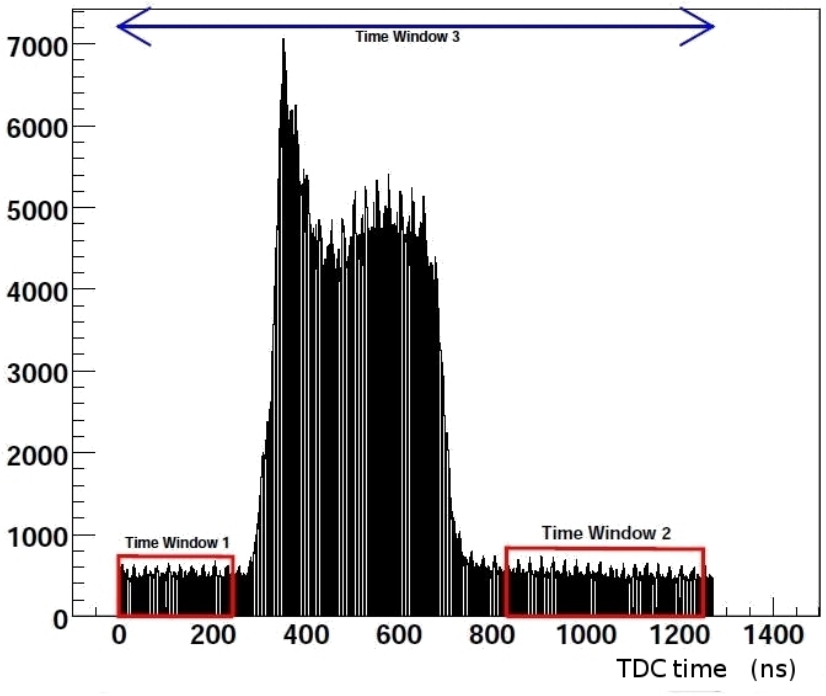


Figure 6.19: The Timebox shows the distribution of the time measured by all the hits recorded by the TDC, in a time window of 1250 ns (Time Window 3, blue). Hits used for noise studies are marked in red squares (Time Window 1 and 2).

Figure 6.20 shows the number of background hits per second and channel by July 2010, when the peak luminosity was still low ($\sim 10^{29}$ cm^{-2}s^{-1}) . On average, background

was 5.40±0.01 Hz per cell. However, the long tail shows a few number of cells very noisy, which means they record a higher number of hits than expected in a normal behaviour. About 19 cells out of near 200 000 which form the DT system, exceeded the threshold of 500 Hz, and about 5 of them exceeded 10^5 Hz. They were mainly localized in the borders of the chamber, which may suggest some interference with the electronics. These cells do not affect the overall DT performance, as they are a tiny fraction of the total cells (less than 0.01%) and they are masked when malfunctioning. The number of masked cells has maintained more or less constant with time during the period under study, during all 2010, although the particular cells did vary. Some noisy cells recovered their normal behaviour, sometimes at the same time of the technical stops of data taking.

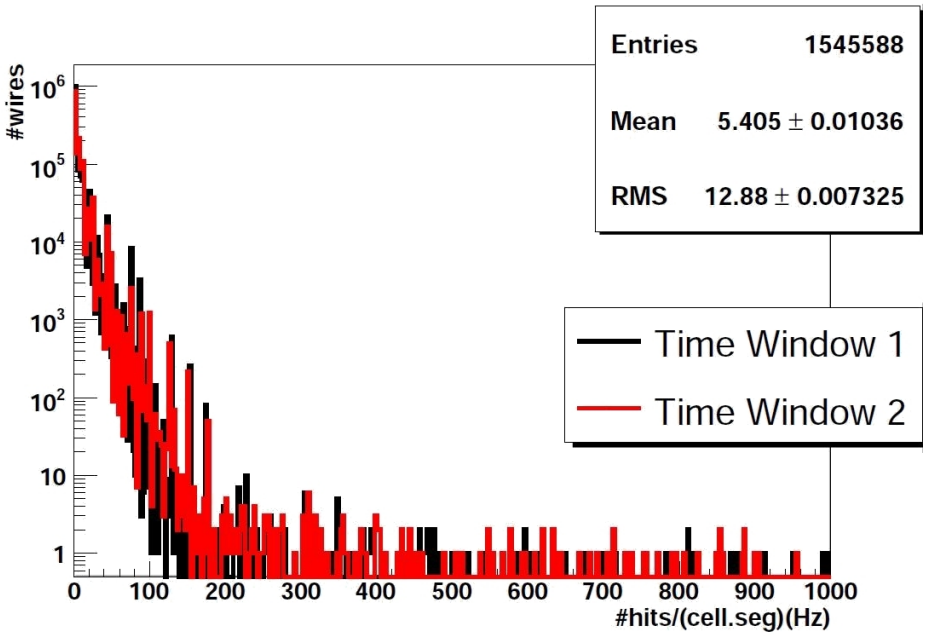


Figure 6.20: Rate of noisy hits per channel for the two subsamples considered.

The other main source of noise were the called "*monster events*" which appeared at the beginning of the LHC running, but disappeared with the increase of the luminosity by the end of 2010. These phenomenon consisted on events where a large fraction of all DT channels received a signal. In particular, the majority of channels in the wheels +1, +2. These events did not affect the muon reconstruction, as "fake" muons were not reconstructed under such conditions, but it increased the size of the events recorded. It affected less than 0.02% of the events recorded.

Both contributions, individual noisy cells and monster events, are present in Fig. 6.21, which shows the dependency of the noise hits with the wheel of the barrel. An asymmetry is observed from the wheels -1,-2 with an average noise per cell of 4 Hz, whereas 0, +1, and +2 show more than 12 Hz. Most of the high value in wheel 0 was due to three very noisy cells that increased the average rate. Masking cells with

more 500 Hz produced the red line, where the rate in wheel 0 was drastically reduced. On the other hand, large contribution to average noise in +1,+2 was removed with a cut in events with more than 500 signals in the DT's (green line). An equivalent quality cut was to make sure that real collisions have taken place, with the advantage that this cut also removed contribution of cosmic muons crossing the detector near the interaction point and therefore being confused with muons from collisions. This was more significant with low luminosity when cosmic background was still significant with respect to actual collisions.

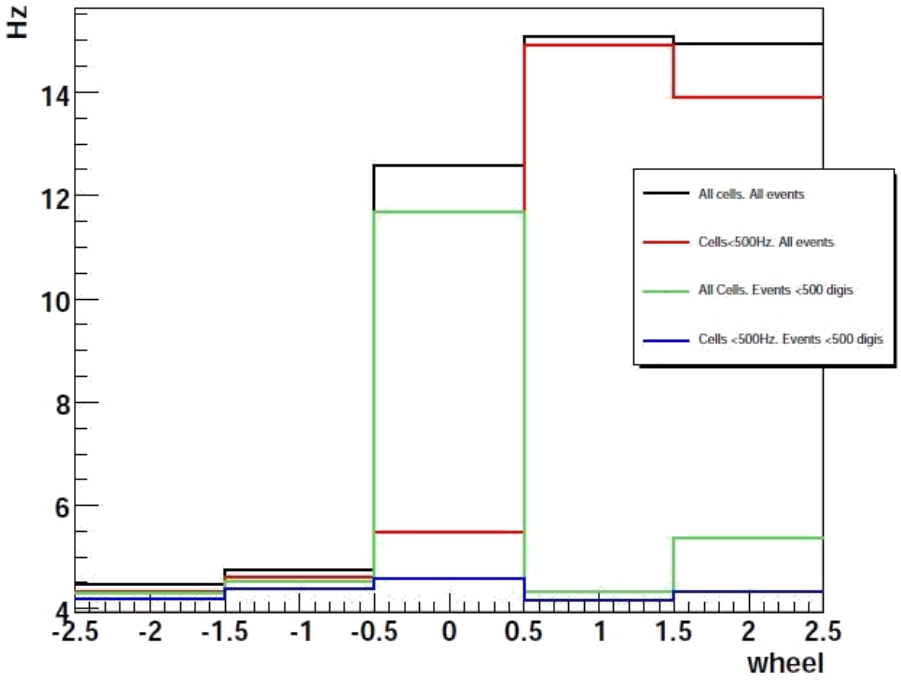


Figure 6.21: Average background noise (Hz) for the different wheels. Different line colours correspond to different cuts imposed to isolate sources of background: no cuts (black), cut on noisy cells (>500 Hz) (red), cut on events with more than 500 hits (green), and combination of both cuts (blue).

Fig.6.22 and Fig. 6.23 show the distribution of the noise in the different sectors before and after the cleaning, respectively. Average noise per cell was reduced to 3.8 Hz/cell homogeneously distributed in all detector.

Influence of the high voltage feeding the anode, cathode and strips has been studied, collecting data with the high voltage off, and no influence was found in the noisy cells nor in the monster events.

The level of noise has been monitored up during all 2010 to study the dependency with the instantaneous luminosity. A linear relationship has been found, rising from 5.4 Hz up to 23.4 Hz for a factor 100 of increase in the instantaneous luminosity. The most affected areas (see Fig. 6.24) are the innermost stations (MB1), suffering

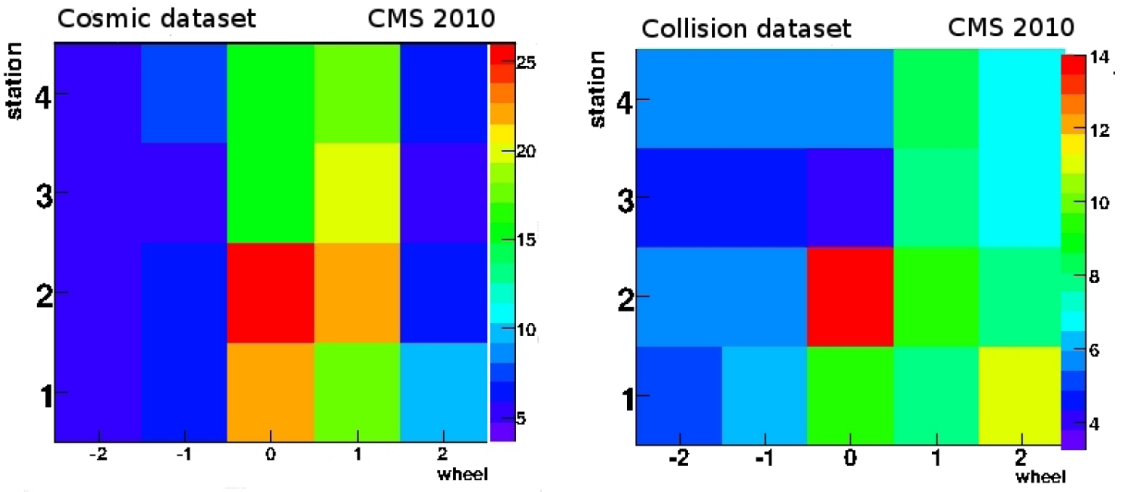


Figure 6.22: Average number of hits per cell and second (Hz/cell) as a function of the station and wheel in the barrel, for a cosmic muon dataset (left) and collision dataset (right). Noise is dominated by a few noisy cells and the left-right asymmetry due to monster events.

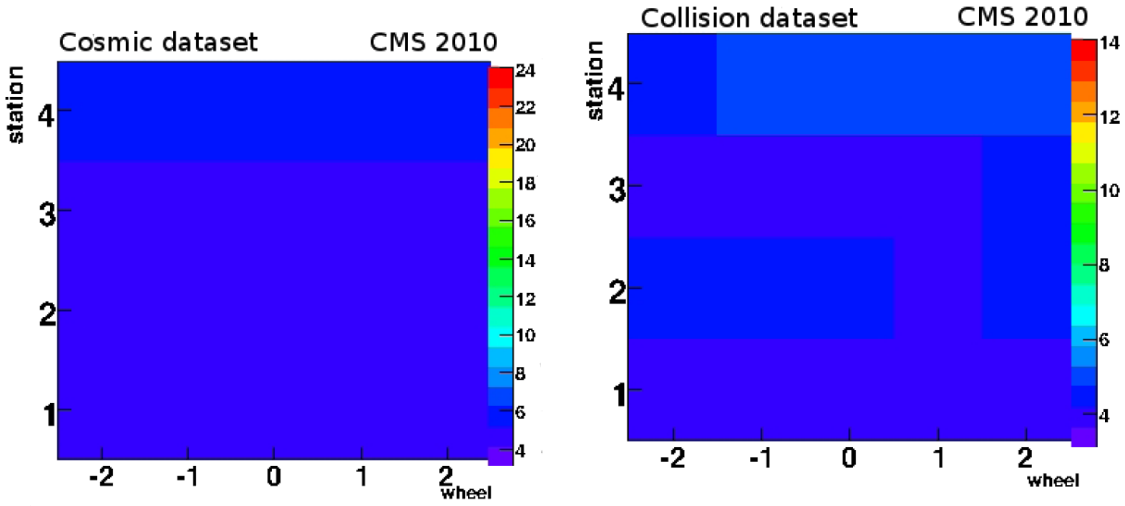


Figure 6.23: Average number of hits per cell and second (Hz/cell) as a function of the station and wheel in the barrel, for a cosmic muon dataset (left) and collision dataset (right), after masking noisy cells with producing than 500 Hz, and filtering events with more than 500 signals in the drift tubes. The average noise is reduced from 5.4 Hz to 3.8 Hz, homogeneously distributed along the detector.

of punch-through radiation. And specially, the external wheels ±2 where the flux of particles from collisions is higher. Moreover, the most external stations (MB4) suffer of the low-neutron radiation existing in the cavern and cosmic flux. An important azimuthal dependency is observed. The upper sectors (4,13), oriented to the surface, exhibit a higher level of noise, whereas the lower sectors, protected by the steel floor and the pedestal of the detector, show lower levels of noise. It is particularly visible the case of the wheel -2, placed under the shaft connecting the cavern with the surface, so suffering a higher flux of cosmic rays.

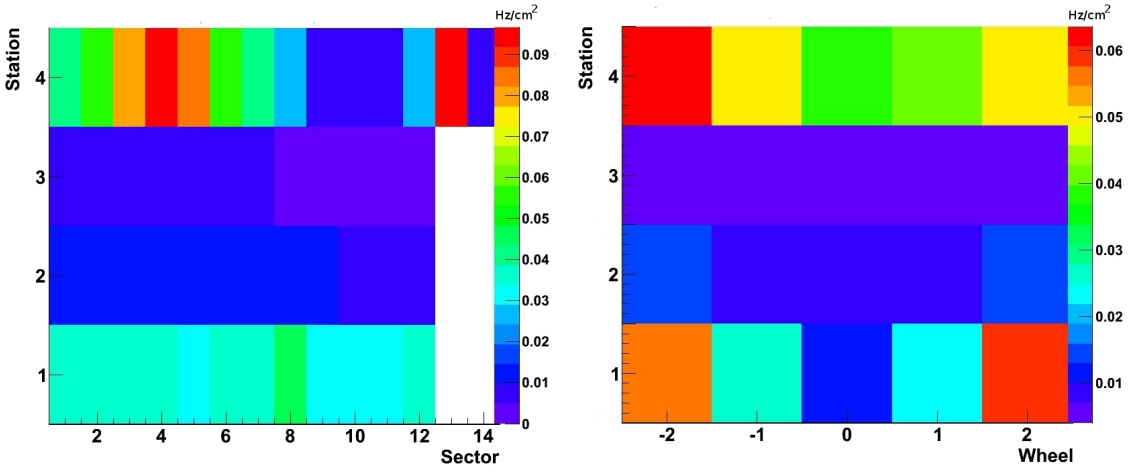


Figure 6.24: Average noise for a instantaneous luminosity of the LHC of 10^{32} cm^{-2}s^{-1}. Left plot shows a station-sector map, averaging over all wheels. Right plot shows the dependency with the wheel and station, averaging over all sectors.

It has been studied the dependency with the time of collisions in the LHC, to dismiss the source of noise as hits produced in collisions at a different time, and consequently assigned a wrong bunch-crossing (BX) in the LHC orbit. For this test, only chambers with no local trigger has been analyzed, as for the ones triggering it is assumed the relation with the collisions.

Figure 6.25 shows the dependency with the BX. From the length of the LHC and the frequency of the radiofrequency caverns, a maximum of 2 808 bunches are allowed. However, not all of them were filled, specially during 2010. Figure 6.25 shows the pattern of noise each 6 BX, matching with the BX when the collisions took place. In conclusion, the out-of-time noise analysed is related to the collisions. However, a small pedestal in the BX previous to the collision is also observed.

Finally, extrapolations to higher luminosities are taken into account. It is important to maintain the level of noise under control and estimate the prospects for a higher luminosity regime of $\sim 10^{34}$ cm^{-2}s^{-1}.

Figure 6.26 shows the expected noise at the nominal luminosity of the LHC. The noise rates induced a high uncertainty in the extrapolated luminosity, so the result is mainly an estimation of the order of magnitude and a confirmation that the low

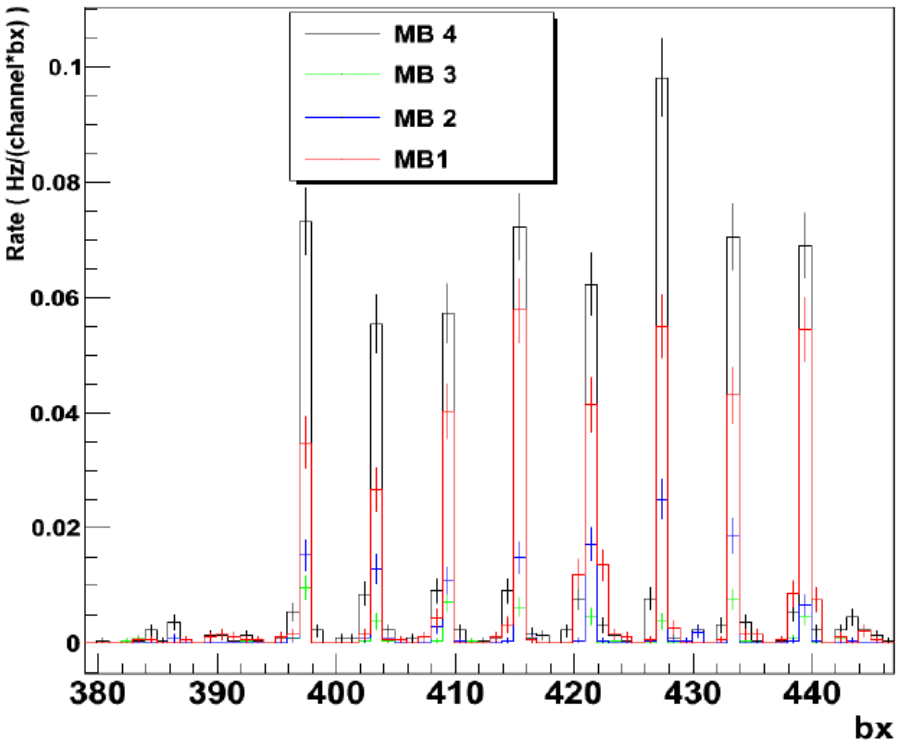


Figure 6.25: Average noise as a function of the LHC orbit bunch-crossing. A zoom is done to see individual bunch-crossings

levels of noise are under control and do not compromise the data taking. It must be mentioned that the extrapolation covers three orders of magnitude in the instantaneous luminosity. Still, very good agreement has been observed comparing to more recent results from 2012 data [74].

To summarize, the drift tube chambers have shown an excellent performance from the very beginning of the LHC running. Efficiency is very high in all parts of the detector, and noise have been understood and kept under control even for higher luminosities without affecting the detection and reconstruction. Improvements in calibration, synchronization and a larger sample have allowed more accurate measurements of single hit resolution.

In conclusion, they have been a fundamental part in the CMS commissioning and in the consequent Physics results.

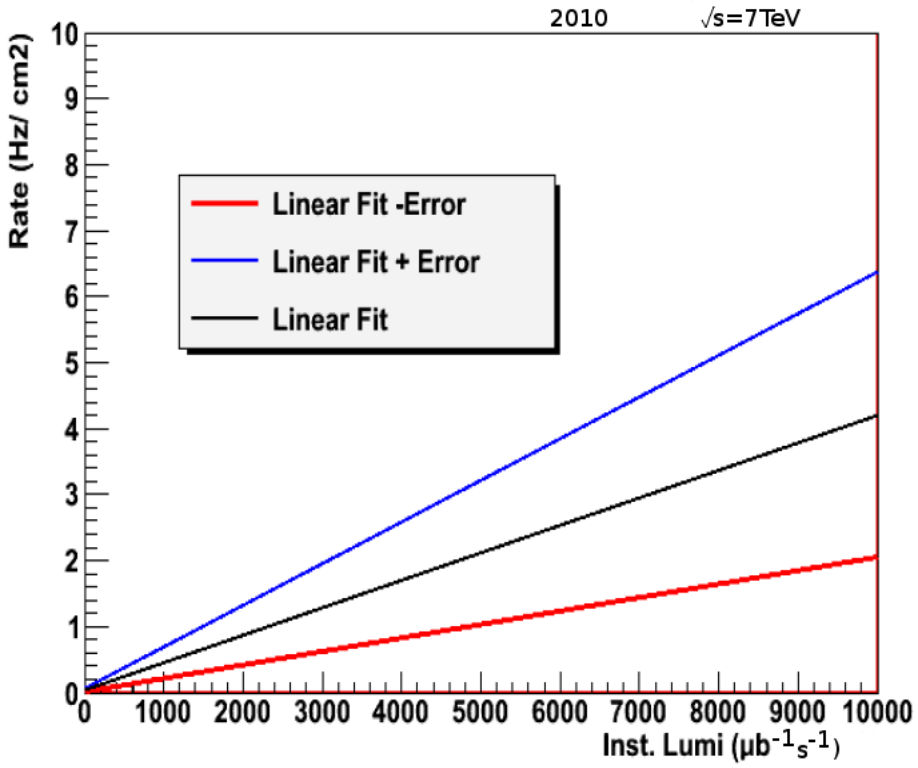


Figure 6.26: Extrapolation of the noise at the expected higher instantaneous luminosities at the LHC

Chapter 7

Lepton and jet reconstruction in the CMS detector

The analyses looking for the $H \to ZZ \to l^+l^-q\bar{q}$ final state involves several reconstructed objects, such as muons, electrons, hadronic jets and derived quantities as the missing transverse energy. This chapter summarizes the algorithms and performance of the triggering, identification and reconstruction processes of these objects, which will be essential in the analyses strategy and the systematic uncertainties determination.

The event reconstruction in CMS is performed via the Particle Flow (PF) algorithm[75, 76], which combines information from all CMS subdetectors to identify and reconstruct individual particles created in the collisions. These particles are classified in several types: charged hadrons, photons, neutral hadrons, electrons and muons. A global event description improves identification of leptons and also the reconstruction of jets and the missing transverse energy.

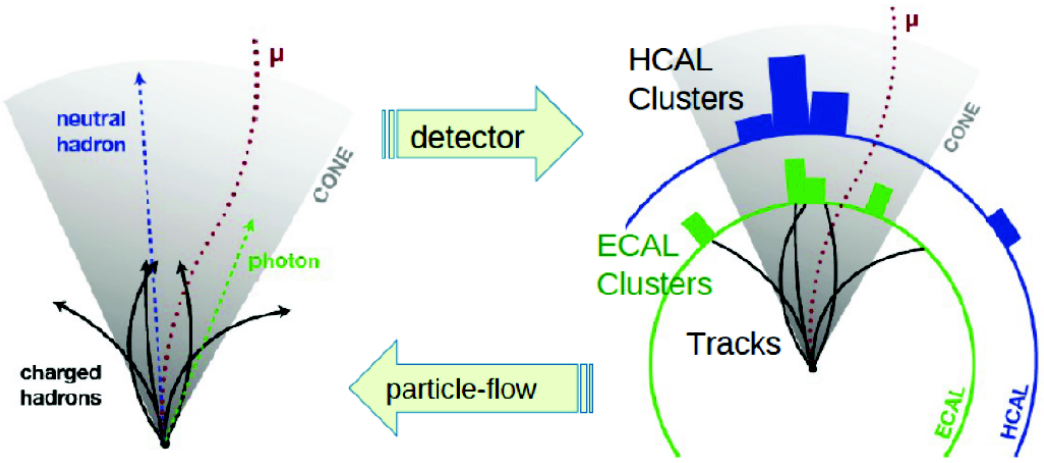


Figure 7.1: Scheme of the particle flow algorithm in CMS.

7.1 Muons

Muon reconstruction is very robust and efficient, attributable to several redundant and accurate layers of detection [67]. As explained in chapter 6, a DT chamber reconstructs a linear segment of the muon track. Similarly, CSC build linear segments from the 6 planes of reconstructed 2-D hits.

Then, segments are extrapolated from one DT or CSC to the next station, using a Kalmar Filter. The magnetic field, the multiple scattering inside the steel yoke, and the energy loses are taking into account. A standalone muon track is reconstructed in the muon system.

The last step of reconstruction is the matching of the standalone track with a track in the inner silicon detector. Two types of muons are considered, according to their reconstruction process:

- Global muon: The standalone track is extrapolated to the tracker and a search is performed in a cone around it to match a tracker track.
- Tracker muon: All tracker tracks with transverse momentum p_T>0.5 GeV are considered as seeds and extrapolated to the muon system, looking for a match with a DT/CSC segment.

A tracker muon is less restrictive in the muon system reconstruction (only requires a muon segment) so it will be slightly more efficient for low-p_T muons, which might not cross enough muon stations as to reconstruct a standalone track. However, over p_T>5 GeV, differences in efficiency are negligible. In fact, more than 99% muons in the detector acceptance are reconstructed by both methods.

Different analyses in CMS may need different commitments between efficiency and purity in the identification of a muon. Several quantities are used to classify identified muons in different types according to their purity: *loose, soft* and *tight* muons. These quantities cover:

- quality of the tracker track: to guarantee a good p_T measurement.
- thresholds in the impact parameter of the track: to suppress muons created in decays in-flight of hadrons (e.g. muons coming from b,c decays) and cosmic rays.
- χ^2 cut of the track fit and requirements in muon system: to further suppress muons from decays in-flight and punch-through.

Table 7.1 summarizes the selection criteria for muons with different level of purity. Loose muon category is skipped as it only requires a muon to be a PF muon with a global or tracker track. Furthermore, an additional selection is dedicated to high-pt muons (over 200 GeV) which is useful in certain analyses.

The purity varies from a 7% of punch-through and fake muon contamination in soft muons, to less than 0.5% in tight muons.

Soft	Tight
Tracker track	global track
-	PF muon
tracker layers w. hits >5	tracker layers w. hits >5
pixel layers w hit >1	Valid pixel hits >0
χ^2/n.d.o.f. of tracker track fit <1.8	χ^2/n.d.o.f. of global track fit <10
Loose IP cuts	Tight IP cuts
-	muon chambers in fit >1
-	muon hits in global track >0

Table 7.1: Requirements on muon identification with different level of efficiency and purity.

The "*tag* & *probe*" method provides an almost unbiased estimation of the efficiency of muon trigger and offline reconstruction. This method uses dimuon resonances (J/ψ to study the low p_T regime, Z for the intermediate one) and *tags* one of the muons with the full selection criteria. The efficiency is evaluated on other muon, (*probe*), subject to different selection criteria depending on the efficiency to measure.

The full efficiency of a reconstructed muon is expressed as the convolution of several efficiencies:

$$\epsilon_\mu = \epsilon_{tracking} \times \epsilon_{RECO+ID} \times \epsilon_{ISO} \times \epsilon_{trigger} \tag{7.1}$$

where $\epsilon_{tracking}$ is the efficiency of the track reconstruction of the inner tracker, $\epsilon_{RECO+ID}$ is the efficiency of reconstructing and identifying a muon, ϵ_{ISO} is the efficiency of the isolation criteria (described below) and $\epsilon_{trigger}$ is the efficiency of the muon trigger system. $\epsilon_{tracking}$ is more than 99% within the tracker acceptance. The $\epsilon_{RECO+ID}$ is (99±0.24%) for soft muons and more than 96% for tight muons (96.4±0.2% in the barrel and 96.0±0.3% in the endcaps), in excellent agreement with simulation.
The muon trigger efficiency is also evaluated with the tag&probe method from dimuon resonance samples. The efficiency depends on the kinematics of the muon and the purity of the identified muon. Most of the inefficiency is due to p_T thresholds in the L1 and HLT algorithms (which affects the very low p_T regime) and due to inefficiency in the barrel-endcap overlap region. For tight muons, a plateau efficiency is reached for p_T over 10 GeV, of $\sim$ 95 % in the central region.

7.1.1 Isolation

The study of the activity of the detector in the vicinity of a muon track allows the discrimination between a muon coming from the decay of a weak boson and and a leptonic decay of an hadron, especially those that contain a heavy-flavour quark. The criteria of isolation is established as a certain threshold on the ratio of the sum of p_T of objects reconstructed in a cone around the muon itself to the muon p_T. This

sum of p_T in the numerator may be based on detector information, using tracker tracks (*tracker isolation*) or tracker tracks and energies measured in ECAL and HCAL towers (*combined isolation*). *PF relative isolation*, in contrast, uses particle-flow reconstructed particles: charged hadrons and transverse energies E_T of all photons and neutral hadrons.

The cone is defined by its radius $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, optimized independently for the different algorithms.

Figure 7.2 shows the efficiency of the isolation criteria as a function of the p_T for the combined tracker+calorimeter algorithm (left) and particle-flow algorithm (right). The thresholds in each case are decided representing the efficiency of the isolation as a function of the value of the threshold, and picking the value where the efficiency reaches a plateau.

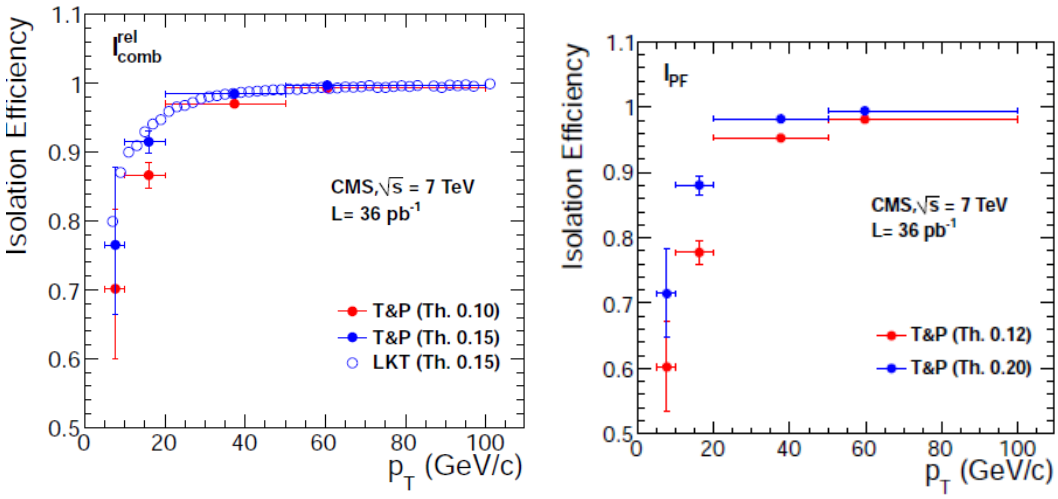


Figure 7.2: Left: efficiency of tracker-plus-calorimeters relative isolation for muons from Z decays as a function of muon p_T. Results corresponding to the threshold values of 0.10 and 0.15 are shown. Right: efficiency of particle-flow relative isolation for muons from Z decays as a function of muon p_T. Results corresponding to the threshold values of 0.12 and 0.20 are shown.

Pileup has a strong effect in the isolation determination. The amount of energy collected in the cone around the muon candidate rises and the isolation decision is biased. The increase of the luminosity in the LHC enforces to develop techniques to remove the pile-up contribution to the amount of energy measured. The pile-up charged particle contribution is easily identified due to the good vertex resolution, but the neutral particle contribution needs further corrections. In the *effective areas* correction, an average pile-up energy density ρ is calculated per event, based on the FastJet reconstruction algorithm [77]. The neutral contribution in the cone is thereby corrected as

$$\sum_{n.c} p_t \rightarrow max\left(0, \sum_{n.c} p_T - \rho \cdot A_{eff}\right) \tag{7.2}$$

where A_{eff} is an effective area of the cone, calculated to remove the dependency on the number of reconstructed primary vertices.
An equivalent method has been developed called *delta beta corrections* ($\Delta\beta$) computes the pile-up charged particle deposits produced in other vertices, and estimates the neutral counterpart from an average charged/neutral particle established ratio.

7.1.2 Momentum scale and resolution

The measurement of the muon transverse momentum is strongly dependent of the alignment of the tracker and the muon system, and the knowledge of the magnetic field and of the energy loss in all the volume of the detector.

Tracker resolution dominates the muon momentum resolution. Only for high-p_T muons, the global muon fit improves the momentum resolution with respect to the tracker fit. The default algorithm for momentum assignment estimates the tracker-only fit momentum and the global-fit momentum. Then, the global fit is chosen when both fits yield a muon p_T above 200 GeV and give the charge-to-momentum ratios q/p that agree to within $2\sigma_{q/p}$ of the tracker-only fit; in all other cases the tracker-only fit is taken.

Muon momentum scale is calibrated using the mass constraint in dimuon decays of J/ψ and Z resonances, at low and intermediate p_T range, respectively. For p_T larger than 100 GeV, cosmic-ray muons are used (except in the region of high pseudorapidity). In this case, the "*Tune P*" algorithm, which takes into account possible radiation losses in the material, is used to improve the p_T resolution.

Muon momentum resolution is also evaluated with J/ψ and Z resonances and cosmic rays, smearing simulation to match data. Resolution depends on the p_T and the region of the detector.

At low and intermediate p_T, the resolution $\sigma(p_T)/p_T$ depends on the p_T of the muon and the region of detection. Figure 7.3 (right plot), shows the dependence of the resolution of a tight muon with η. The dependence on p_T, on average, varies from 1.8% for at p_T=30 GeV to 2.3% at p_T=50 GeV.

Resolution q/p_T for high-p_T muons, over 200 GeV, measured with cosmic rays, is shown in Fig.7.3 (left plot), for different algorithms.

In conclusion, the specifications of $\sigma(p_T)/p_T \sim 1\%$ at 100 GeV and $\sigma(p_T)/p_T \sim 10\%$ at 1 TeV are satisfied.

7.2 Electrons

The reconstruction of electrons is done together by the ECAL and the tracker detector. One critical aspect is the big amount of silicon material the electrons have to traverse, which induces big losses through bremsstrahlung photon radiation. High energy electromagnetic showers spread laterally over several crystals, so the reconstruction starts with "hot" crystals in the ECAL being grouped in clusters. To

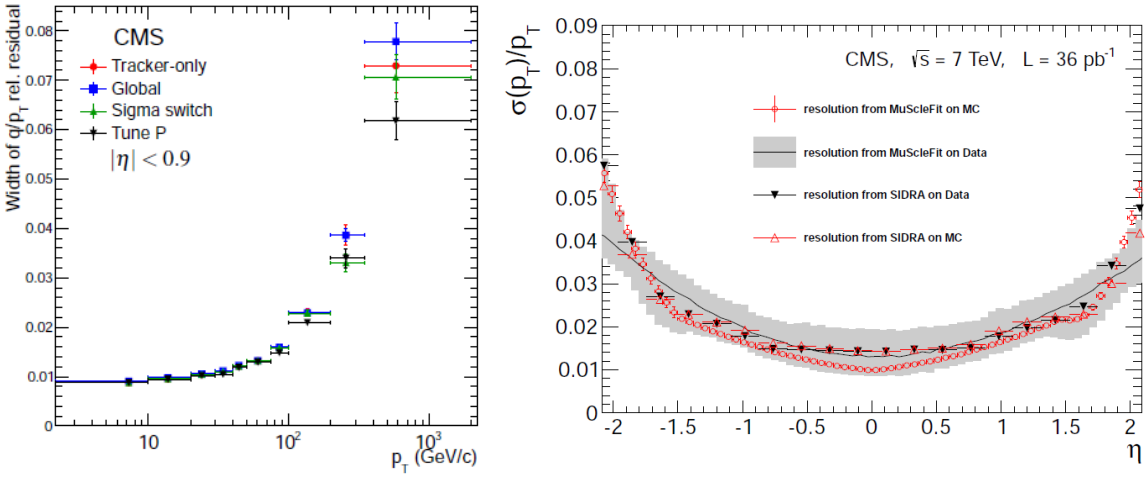


Figure 7.3: Left: Widths of Gaussian fits of the distributions of the muon q/p_T relative residuals as a function of the p_T of the muon, for different reconstruction algorithms: tracker-only and global fits, and the output of the sigma-switch and Tune P algorithms. Right: Relative transverse momentum resolution $\sigma(p_T)/p_T$ in data and simulation for muons of $p_T \sim$ 50 GeV.

recover radiation from bremsstrahlung losses and conversion of photons from electrons, clusters are grouped along the ϕ coordinate in superclusters. Because of the different geometry of the detector in barrel and endcap, different clustering algorithms are used in different regions [78]. Superclusters are used as seeds to find hits in the pixel detector. The track is built with a Kalman Filter. The Gaussian Sum Fitter (GSF) algorithm is used to refit and re-estimate the inner track (reconstructed with the tracker) taking into account the bremsstrahlung losses.

The final identification is based in different detector variables, which may be used independently with optimized thresholds for each one (cut-based identification), or combined with some multi-variate analyses technique (MVA identification).

- $\Delta\eta(SC, \vec{p}(vtx))$: η separation between the ECAL supercluster position and the track direction in the vertex, extrapolated to the ECAL, assuming no radiation.

- $\Delta\phi(SC, \vec{p}(vtx))$: ϕ separation between the ECAL supercluster position and the track direction in the vertex, extrapolated to the ECAL, assuming no radiation.

- E_{HCAL}/E_{ECAL}: ratio of energy collected in the HCAL tower behind the ECAL weighted position of the ECAL supercluster, to the ECAL supercluster deposit, used to suppress hadronic jets misidentified as electrons.

- $\sigma_{i\eta,i\eta}$: width of the ECAL supercluster along the direction computed for all the crystals in the 5x5 blocks of crystals centred on the highest energy crystal of the seed supercluster.

- d_{xy},d_z: transverse and longitudinal impact parameter of the track, with respect to the reconstructed vertex.

- $1/E_{tot} - 1/p$: E is the energy measured in the ECAL and p is the momentum measured in the tracker at the vertex position.

- The number of missing hits in the back-propagation of the track to the beam line.

Reconstructed electrons from converted photons in the tracker are a non-negligible background to "prompt" electrons at hadron collisions (e.g. electrons originated in the primary interaction point, like those from W and Z boson decays). Therefore, efficient methods to identify and reject such electrons are fundamental to any analyses that involves "prompt" electrons, such as check missing expected hits in front of the innermost tracker layers.

Different working points are established, as for the muons, depending on the commitment between efficiency and purity of the identification [79]. Table 7.2 shows the full identification criteria for the Loose Working Point, representing an overall efficiency above 90%. Figure 7.4 shows the efficiency as a function of the p_T for the Medium Working Point, which has tighter requirements, in two different acceptance areas.

The electron isolation is performed as explained in section 7.1 for muons. Effective area corrections are also applied to correct pile-up contributions.

	Endcaps	Barrel
$\Delta\eta(SC, \vec{p}(vtx))$	<0.009	<0.007
$\Delta\phi(SC, \vec{p}(vtx))$	<0.1	<0.15
$\sigma_{i\eta,i\eta}$	<0.03	<0.01
E_{HCAL}/E_{ECAL}	<0.1	<0.12
d_{xy} (cms)	<0.02	<0.02
d_z (cm)	<0.2	<0.2
$1/E_{tot} - 1/p$	<0.05	<0.05
Missing Hits	<=1	<=1
Conversion vertex fit prob.	$<10^{-6}$	$<10^{-6}$

Table 7.2: Electron identification cuts for the Loose Working Point.

As for the case of muons, the electron energy scale is calibrated using the tag&probe method using $Z \to e^-e^+$samples. Additional corrections or smearings are applied to the simulation to account for possible differences with respect to the data reconstruction. The energy resolution is also evaluated with this tag&probe method, splitting the sample according to the amount of bremsstrahlung losses, in order to exploit calorimetric or tracker variables. Furthermore, it is significantly improved using multivariate regression techniques.

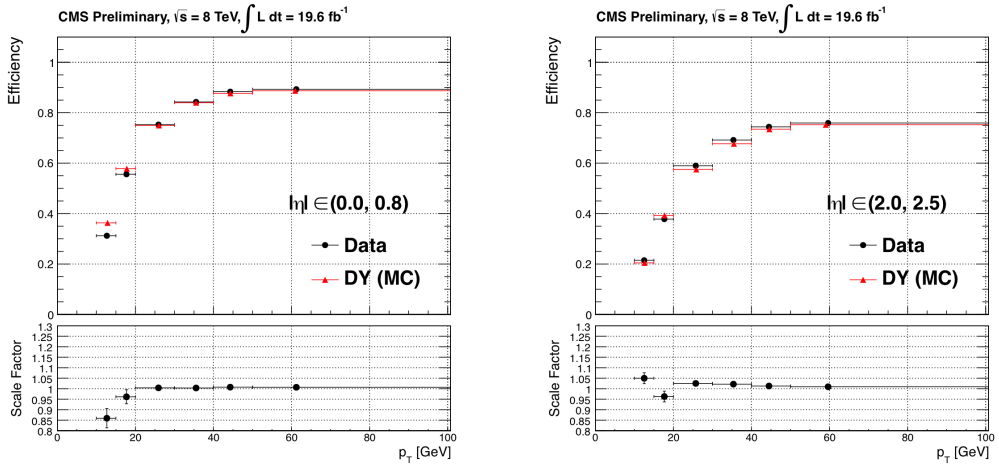


Figure 7.4: Electron selection efficiency for the medium working point (WP) on data and on a Drell-Yan Monte Carlo simulation sample as a function of the electron pT, for two η regions: $0 < \eta < 0.8$ (left) and $2.0 < \eta < 2.5$ (right). Only statistical errors are shown.

Figure 7.5 shows the measured resolution in data an simulation for different categories in $Z \rightarrow e^{-}e^{+}$resonances [79]. The resolution in the barrel is in 1-2.5%, whereas in the endcaps rises to near 3-4%, depending on the category.

7.3 Jets

Hadronic jets are the result of the hadronization of quarks. They are typically composed by 65% charged hadrons, 25% photons, 10% neutral hadrons. Consequently, about a 90% of energy is determined by tracker+ECAL with high precision. For this reason, the particle flow algorithm involving the whole detector reproduces better the simulation in different kinematic variables, like p_T or direction, which leads to a better missing energy determination. Furthermore, the particle-flow algorithm will identify different components inside the jet, contrary to a calorimetric-only reconstruction (called *calo-jets*).

The PF algorithm will identify candidates of muons, electrons (as explained above), photons, and charged and neutral hadrons. Photons correspond to ECAL deposits not compatible with a tracker track. Charged hadrons correspond to HCAL and/or ECAL deposits matched to a inner track and not compatible with an electron, whereas neutral hadrons are identified as HCAL deposits not matched to any track.

The formation of the PF jet will, in consequence, depend only on the clustering algorithm of the PF candidates. Different algorithms can be used for clustering one jet, although the most extended in CMS is the *anti-kt* algorithm [80], which creates perfectly cone-shaped jets of radius R. The distance of a object i to the beam and to another object j is defined as

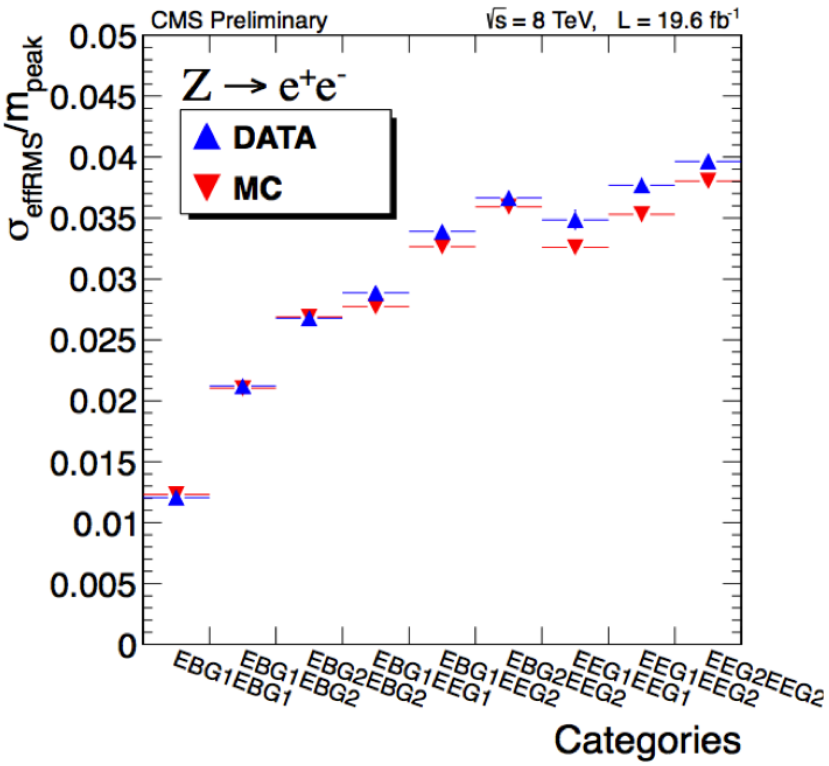


Figure 7.5: Instrumental di-electron mass resolution as measured from $Z \to e^- e^+$events and compared to simulation. Events are categorized according to the electron class and pseudorapidity region of each leg (G1: electron is golden or bigbremsstrahlung, G2: electron is showering or crack or bad-track, EB: electron is in ECAL barrel, EE: electron is in ECAL endcaps).

$$d_i = \frac{1}{p_{T_i}^2} \tag{7.3}$$

$$d_{ij} = min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right)\frac{\Delta R_{i,j}}{R^2} \tag{7.4}$$

where p_{T_i}, p_{T_j} are the two particles transverse momentum, $R_{i,j} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the Euclidean distance between them in the $\eta - \phi$ plane and R is the radius of the cone, a parameter of the algorithm. It will successively take a decision according to the minimum of these two quantities, for all the reconstructed particles (in this case PF candidates). If $d_i < d_{ij}$, the i-th object will be selected as a jet itself and removed from the list. If $d_{ij} < d_i$, both objects will be merged into one single object by adding their quadrimomenta. This iterative process privileges the clustering of low $p_T$particles around a high p_T seed, instead of several low p_T particles clustering by themselves. This has several advantages from the experimental point of view, as it simplifies corrections from pile-up and the event interpretation. The analyses presented in these thesis have

been performed using the *anti-kt* algorithm with a radius of R=0.5.

One second important step is the correct conversion of the measured energy deposits to the real p_T or energy of the original particle. The non-linear response of the detector implies a hard work on jet calibration. Several corrections are considered sequentially based on simulation and recorded data.

- L1 correction removes the average flow of energy coming from pile-up contribution, through the FastJet algorithm [77].
- L2 *relative* correction improves the dependence on the pseudorapidity variable.
- L3 *absolute* correction corrects the dependence on p_T so, at the end, the average jet matches the average generator level particle jet
- Additionally, flavour corrections are made to take into account dependencies in the flavour content of the original parton.

Once the response of the detector is uniform, an absolute jet energy correction factor (called jet energy scale) is calculated using a sample of jet+γ events [81]. They must be balanced in the transverse plane to the beam, therefore an accurate measurement of the photon energy is used to calibrate the jet energy.

Figure 7.6 shows the energy correction uncertainties for jets of p_T=30 GeV. The overall uncertainty is better than 3% in the barrel, and better than 5% in the endcaps.

The jet energy resolution is also measured with the same method, both with Z+jets or γ+jets samples [81]. Figure 7.7 shows the jet energy resolution for two η regions. It is higher than $\sim$ 10% only for jets with p_T<100 GeV.

7.3.1 B-hadron identification

The identification of jets originated from b quarks is of great important in CMS. Many processes, like top or Higgs decays, involve b-quarks, and a good bottom identification is a powerful tool to discriminate signal from background events. Different techniques, called *b-tagging* algorithms, are devoted to identification of b-jets with the highest efficiency and purity possible [82].

They rely on different properties of the B hadrons, like high mass, long lifetime or hard fragmentation:

- The *Track Counting* algorithm exploits its long lifetime. It calculates the signed impact parameter (IP) significance of all good tracks, and orders them by decreasing significance. Its b-tag discriminator is defined as the significance of the N'th track. Two cases are considered for N = 2 (high efficiency) or N = 3 (high purity).
- The *Jet Probability* algorithm also exploits the long lifetime to estimate the likelihood that all tracks associated to the jet come from the primary vertex,

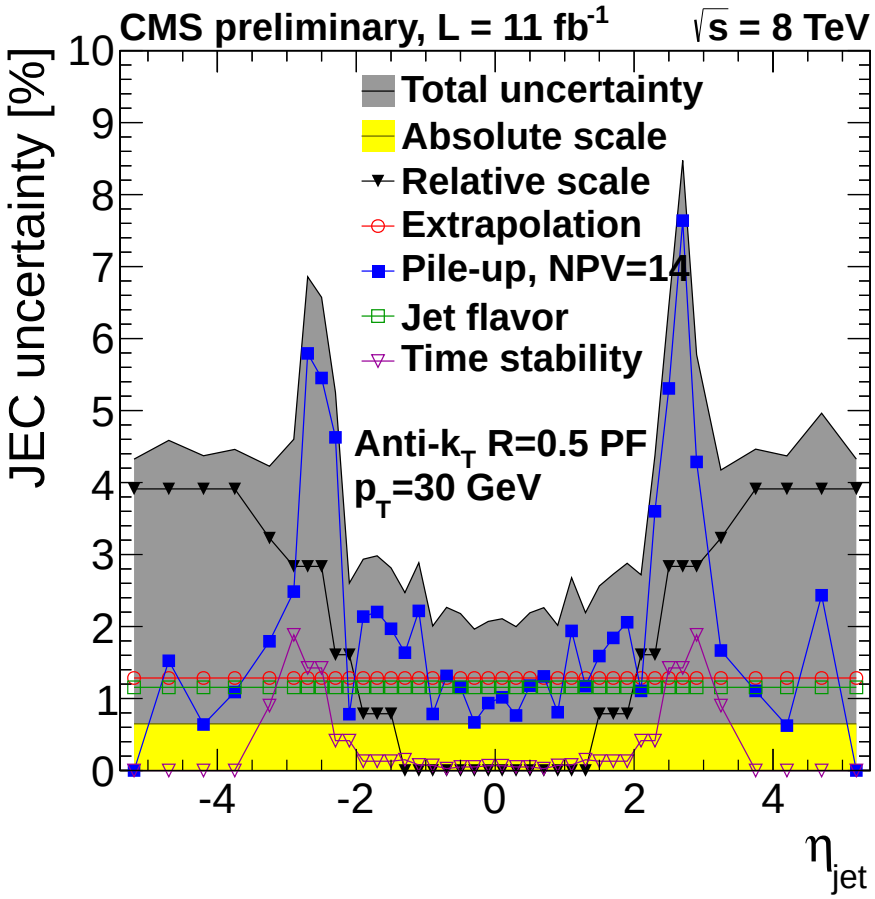


Figure 7.6: Jet energy correction uncertainties as a function of jet pseudorapidity for jets with transverse momentum equal to 30 GeV. Different contributions are shown with markers of different colors, and the total uncertainty is shown with a grey band.

using the inverse as discriminator.

$$P_{bjet} = \Pi \sum_{i=0}^{N-1} \frac{(-ln\Pi)^i}{i!} \qquad \Pi = \sum_{i=0}^{N-1} P_{track}(i) \tag{7.5}$$

where N is the number of tracks considered in the jet and P_i is the compatibility of track "i" with the primary vertex, based on the probability density functions for the IP significance of tracks.

- The *combined secondary vertex* method exploits all known variables which can distinguish b from non-b jets, combining information about impact parameter significance, the secondary vertex and jet kinematics.

- The *soft muon/electron* technique search for the lepton from a semi-leptonic B decay, which typically has a large p_T^{rel} with respect to the jet axis.

Figure 7.8 shows the efficiency curves of the b-identification compared to the light flavour or gluon misidentification of the jet. Minimum thresholds on these

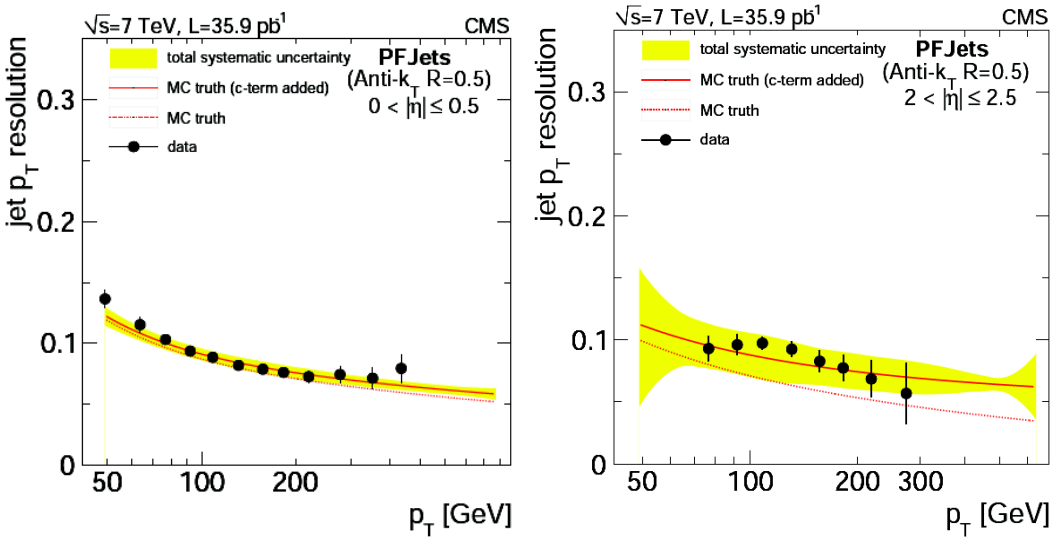


Figure 7.7: Bias-corrected data measurements, compared to the generator-level MC (denoted as MC-truth) pT resolution before (red-dashed line) and after correction for the measured discrepancy between data and simulation (red-solid line) for PF jets in $|\eta|$ <0.5 (left) and 2.0< $|\eta|$ <2.5 (right)

discriminators define several working points with a nominal probability of light flavour jet contamination, established at an average jet p_T of ∼80 GeV. The "loose" working point corresponds to a 10% misidentification, "medium" to a 1% and the "tight" WP to a 0.1%

7.4 Missing Energy

CMS is a full coverage hermetic detector which identifies and reconstructs particles in almost all solid angle. However, the final reconstructed energy may be unbalanced with respect the initial energy. The presence of *missing energy* in the event reconstruction may be due to different (non-exclusive) causes:

- the presence of neutrinos, or other weakly-interacting particle, in the final state, undetectable for CMS.
- the incomplete acceptance of the detector. For example, particles may be created in a very forward direction (close to the beam pipe) where they cannot be detected.
- energy resolution effects in the event reconstruction.

This means that, even if there are no real neutrinos in the event, some missing energy still may be present. The missing energy in the transverse plane (*missing transverse energy*, $\not{E}_{\mathrm{T}}$) is a very important variable in many analyses. Different

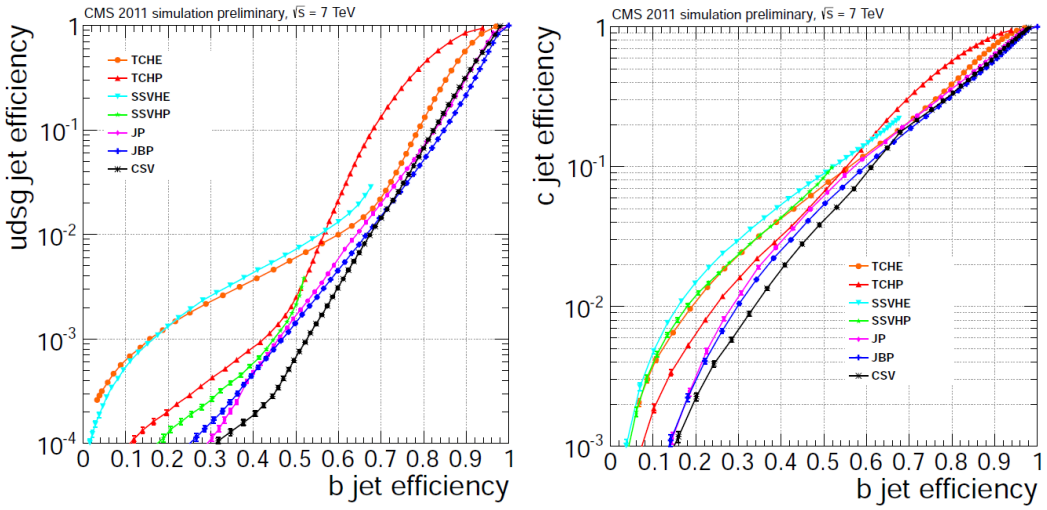


Figure 7.8: Performance curves obtained from simulation for different algorithms: light flavour (left) and c (right) efficiencies as a function of the b efficiency.

estimations of the $\not\!\!E_{\mathrm{T}}$ are performed, in the same sense as the jet reconstruction: using only calorimeter information, including also tracker information, or the particle-flow algorithm. The PF $\not\!\!E_{\mathrm{T}}$ is calculated as the opposite of the sum of quadrimomenta of all reconstructed candidates. Figure 7.9 shows the reconstructed $\not\!\!E_{\mathrm{T}}$ in two different periods. A low-pile-up scenario in the first part of 2011, and the higher luminosity scenario of 2012.

Another useful variable to discriminate the existence of "real" $\not\!\!E_{\mathrm{T}}$ (i.e., due to neutrino production) is the $\not\!\!E_{\mathrm{T}}$ significance, based the ratio of the likelihood that an observed $\not\!\!E_{\mathrm{T}}$is consistent with a real $\not\!\!E_{\mathrm{T}}$, to the likelihood that it is consistent with a fluctuation from zero because of detector-related limitations like finite measurement resolution:

$$\lambda = 2ln\left(\frac{\mathcal{L}(\vec{\varepsilon} = \vec{\varepsilon_T})}{\mathcal{L}(\vec{\varepsilon} = 0)}\right) \tag{7.6}$$

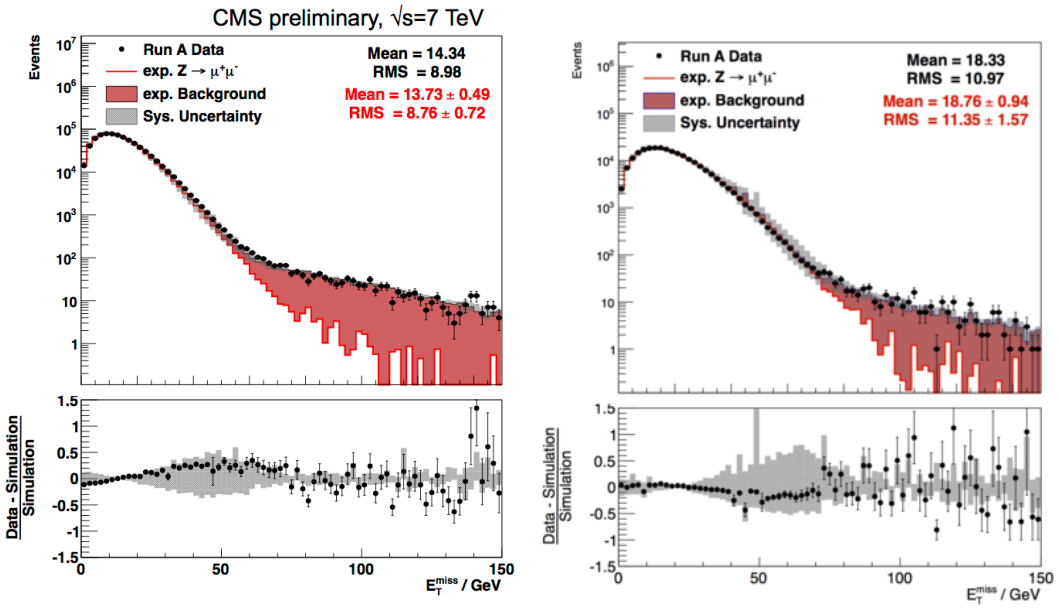


Figure 7.9: Missing transverse energy $\not\!\!E_{\mathrm{T}}$ for two different running conditions: run 2011A (left) and run 2012 (right).

Chapter 8

Search for the Higgs boson in the $H \to ZZ \to l^+l^-q\bar{q}$ decay channel in pp collisions at $\sqrt{s}$= 8 TeV

Chapter 3 explained the mass mechanism and its relevance in Physics. In particular, the SM Higgs boson decays were described according to its mass hypothesis, in section 3.2.2. Huge effort has been put by the CMS Collaboration to investigate all possible decay channels and final states, not only because of its individual importance, but also to achieve the best possible sensitivity by combining results all channels. Special interest had the region beyond $m_H \sim 2m_Z$, where the Higgs essentially decays into WW or ZZ [6]. It had not been explored by Tevatron nor LEP, and it would imply new Physics beyond the Standard Model at energies below the Grand Unification Scale.

The $H \to ZZ \to l^+l^-q\bar{q}$ was a very promising final state for Higgs searches for several reasons:

- The full invariant mass of the candidate is reconstructed, as there are no neutrinos in the final state, as it happens in final states with one $W^{\pm} \to l^{\pm}\nu$ or $Z \to \nu\nu$ decay.

- Most of the Z boson decays are into quarks ($\sim$70% of the total [1]) so the branching fraction of $H \to ZZ \to l^+l^-q\bar{q}$ is much higher compared to other ZZ decays (BR(ZZ$\to$2l2q) = 20 BR(ZZ$\to$4l) = 3.5 BR(ZZ$\to$2l2ν), as it was depicted in Fig. 3.5 in chapter 3).

Because of the W and Z boson branching ratios, the four jet final state would be the most frequent Higgs decay by far. However, the rate of jets produced in the LHC is so huge, that the Higgs signal is indistinguishable from other SM QCD processes. In this scenario, the $H \to ZZ \to l^+l^-q\bar{q}$ offered a balance between higher rates and lower background. This feature was more relevant at larger mass hypothesis, where the background rates diminished and these channel even improved the sensitivity of golden channels like the four lepton final state, near $m_H \sim$1 TeV.

Even after the discovery of a Higgs boson in 2012 reported in section 3.3, Higgs-like signatures are very interesting [47]. Many models predict the existence of additional Higgsses or Higgs-like resonances at higher mass, like Two Higgs Doublet Models (2HDM) or Electroweak singlets coupled to the Higgs boson [83]. WW and ZZ decays still prevail at these mass ranges, so the final state of two leptons and two jets maintains its importance.

Two separate, although similar, analyses have been performed with the data collected during 2011 from collisions at $\sqrt{s}$ =7 TeV [4] in the center of mass, and during 2012 with the data from collisions at $\sqrt{s}$ =8 TeV [5]. I took part on both of them with relevant contributions.

Due to the similar strategies adopted in both analyses, only the analysis performed with data from 2012 will be described in detail in the present chapter. In chapter 9 the main features of the 2011 analysis and its differences with respect to the 2012 one are presented.

Finally, chapter 10 presents the final results and its statistical interpretation under the SM Higgs hypothesis. No other interpretations beyond Standard Model have been considered for this thesis, keeping the SM as a good benchmark for any of them. Specifically, different production rates than the ones expected in the SM, including the existence of resonances in any region of the m_{ZZ} spectrum, may be a clear indication of New Physics.

8.1 Overview of the $H \to ZZ \to l^+l^-q\bar{q}$ channel

The $H \to ZZ \to l^+l^-q\bar{q}$ analysis looks for events with two leptons of the same flavour and opposite charge, which may be electrons or muons ($e^+e^-, \mu^+\mu^-$), and resonant at the Z mass, and two jets also resonant at the Z mass, as represented in Fig 8.1. Figure 8.2 in turn shows a real event display of a $H \to ZZ \to l^+l^-q\bar{q}$ candidate in the muon channel, recorded in the CMS detector.

Several SM processes lead to a similar signature. The main background is the production of a Z boson in association with at least two jets produced via strong interactions (commonly known as Z+jets background). These events look very similar to the signal signature and are very difficult to discern. On top of that, they are produced 10^5 times more frequently than the $H \to ZZ \to l^+l^-q\bar{q}$ rate.

In order to maximize the sensitivity of this channel, the sample is splitted in six exclusive channels. On one hand, electron and muon channels are considered separately. On the other hand, a b-tagging algorithm is applied to the jets, and the sample is splitted into three categories according to the number of b-tagged jets: 0, 1 or 2. The $H \to ZZ \to llb\bar{b}$ final state has much better signal over background ratio S/B, because most of background jets are light jets, whereas a 15% of the Z decays are $Z \to b\bar{b}$. Unfortunately, the size of the subsample is much lower. On the other extreme, the 0-btag category presents much larger statistics, but with less sensitivity. Combining all six categories maximizes the outcome of the result.

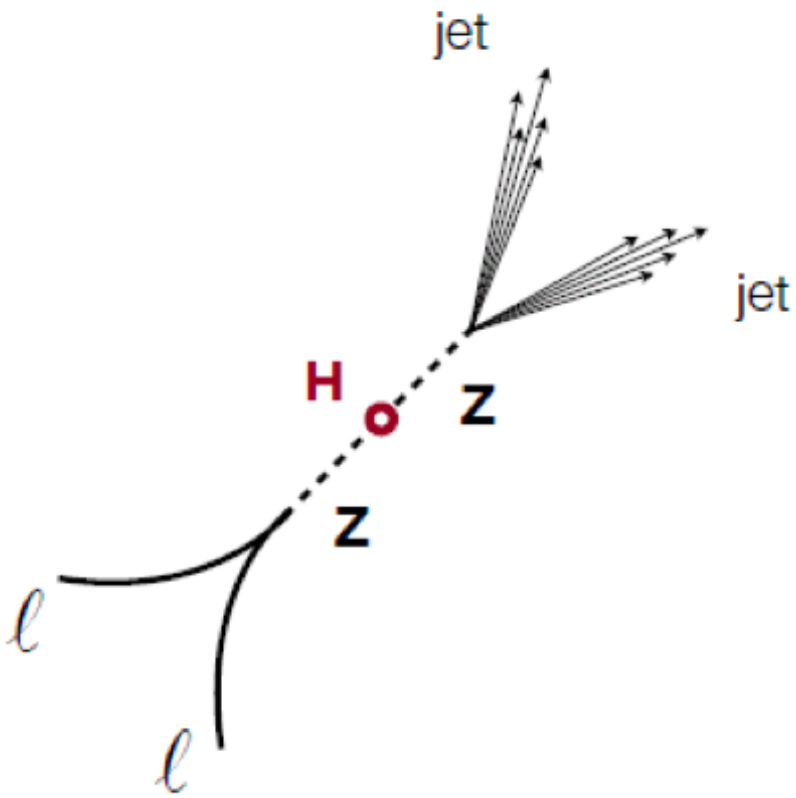


Figure 8.1: Schematic drawing of the $H \to ZZ \to l^+l^-q\bar{q}$ decay.

The second main background is the $t\bar{t}$ production, in its dilepton channel:

$$t\bar{t} \to W^+b\, W^-\bar{b} \to l^+\nu_l l^-\nu_l b\bar{b} \tag{8.1}$$

This background is more important in the 2-btag category, as more than 95% of the times the top quark decay into b quark. Fortunately, an important fraction of this contribution is removed from the kinematics of the two leptons, because they are not resonant at the Z mass.

Finally, the production of dibosons ZZ, WZ and, to lesser extend, WW must be considered. These events are topologically almost identical to the Higgs decay, but fortunately the rate of production of dibosons is very low.

Globally, the background cross section largely exceeds the signal rates. However, different strategies are developed to filter as much as possible to increase the S/B ratio, and will be described below.

As there is no missing transverse energy in this final state, the full invariant mass of the 4-object is reconstructed. It is used as discriminant to look for excesses of events and resonances.

The interpretation of the results is based on a statistical hypothesis test. In order to avoid any possible bias from the research team in the decision of the analysis, a blind

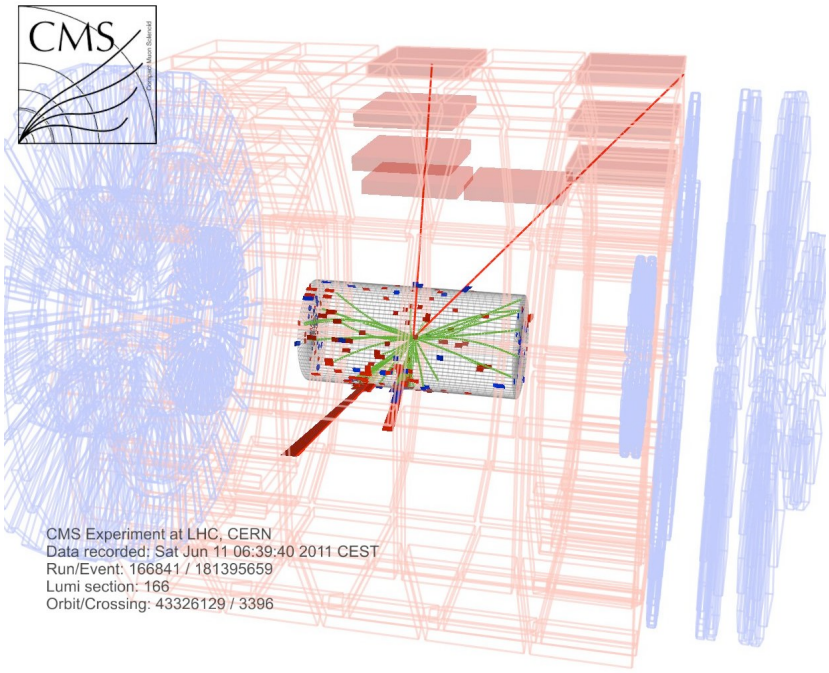


Figure 8.2: Event display of a $H \to ZZ \to l^+l^-q\bar{q}$ candidate recorded in the CMS detector.

policy is established on the recorded data. All the analysis strategy and the selection requirements are decided in a control region, where the kinematics of the background is similar but the signal is expected to be negligible, whereas the optimized subsample for the Higgs search, called "signal region", remains blind. Only after all the analysis steps are validated, the signal region in the data is analyzed.

8.2 Dataset and trigger

8.2.1 Dataset

The present analysis has been performed with the full dataset recorded during 2012 of collisions at $\sqrt{s}$= 8 TeV, adding up an integrated luminosity of $\mathcal{L}$=19.6±0.8 fb^{-1}, four times the luminosity recorded at $\sqrt{s}$= 7 TeV. The data are packed into different primary datasets, centrally produced by CMS for all the collaboration, according to the signatures of particles and jets (physics objects) identified by the high level trigger (HLT). Only data passing quality criteria and certified as reconstructed in good conditions for all subdetectors are used in this analysis. Primary datasets containing at least two electrons or two muons are selected: *DoubleMu* and *DoubleElectron* datasets. Additionally, a *MuEG* dataset containing at least one electron and one muon is used for the $t\bar{t}$ background determination, as described in section 8.4.

Official samples of simulated events generated with Monte Carlo (MC) techniques

have been produced by CMS to characterize background and signal processes.

For the Drell-Yann background, both inclusive *Z+Jets* samples and parton-exclusive *Z+n jets* have been used, where n=1,2,3 and $\geq$4, and the Z decays leptonically. They are produced with the MadGraph V5 1.3.30 [84] element generator. Parton distribution functions (PDF) are modeled by using the parameterizations at next-to-next-to-leading-order (NNLO), and a threshold in the mass of the dilepton pair is imposed at generator level ($M_{ll} > 50$ GeV). The $t\bar{t}$ sample used to study the $l\bar{l}b\bar{b}$ final state uses POWHEG generator [85, 86, 87] at next-to-leading order (NLO). Inclusive SM diboson production WW, WZ and ZZ, is simulated with PYTHIA [88], using normalizations provided by NLO calculations [89].

For the simulated signals, a set of samples of the $H \to ZZ \to l^+l^-q\bar{q}$ process for both gluon fusion and vector boson fusion production, have been simulated for different Higgs masses, using POWHEG. Parton distribution functions (PDF) are modeled using the parameterizations CT10 at NLO [90] and CTEQ6 [91] at LO. At generator level, events are weighted according to the total cross section $\sigma(pp \to H)$, which contains contributions from gluon fusion computed to next-to-next-to-leading order (NNLO) and next-to-next-to-leading-log (NNLL), and from weak-boson fusion computed at [6]. This simulation includes the spin correlations for the Higgs decay chain, which will be useful in the event selection. The $H \to ZZ$ cross sections and branching fractions are provided by the Higgs Cross Section Working Group [92], whereas the Z decays are taken from the PDG [1].

In all cases, the GEANT package [71] has been used to simulate the radiation-matter interaction with the detector.

A full list of data and MC samples with their sizes is found in Appendix A.

8.2.2 Pile-up reweighting

The MC samples are designed to reproduce the conditions of the LHC, including the number of pile-up and its consequences in the reconstruction process. However, the real conditions of the LHC have changed a lot through time, in particular the peak luminosity. This implies that different data taking periods have different number of pile-up, and the MC does not reproduce correctly the data.

CMS handles this by introducing a weight to the simulated events, based on the ratio of the number of pileup measured in data to the number of pileup generated in the MC. Pileup in data is measured from the instantaneous luminosity measured by CMS and the total inelastic proton-proton cross section.

Figure 8.3 compares the number of interactions in data recorded in 2012 and in the MC simulation, for the last reprocessing of the events used in this analysis. Left plot shows a comparison of normalized distributions before the reweighting, whereas the right plot shows the number of reconstructed vertices after the weighting, for all the processes involved.

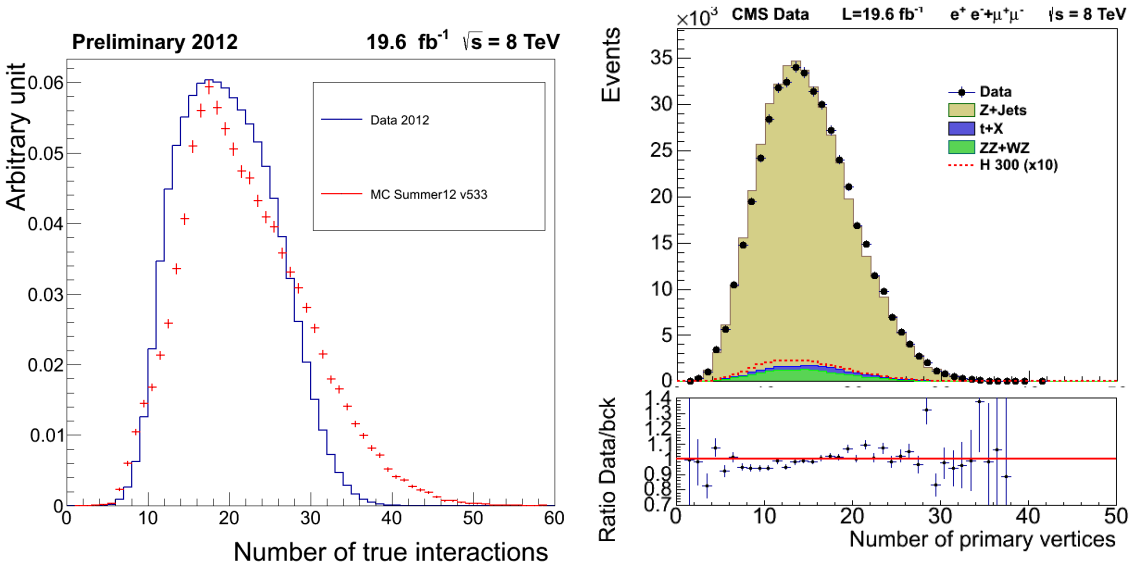


Figure 8.3: Left: number of true interactions for the 2012 data, and for the MonteCarlo production of Summer12 reprocessed with the CMSSW 5.3.3 version. Areas are normalized to one. Right: Number of reconstructed vertices in data and MC after the pile-up reweighting of the simulation. Areas are normalized to the cross section of the corresponding process.

8.2.3 Trigger

Only events satisfying the lowest-threshold unprescaled trigger paths are considered, called:

- For double muons:
  - HLT_Mu17_Mu8
- For double electrons:
  - $HLT\_Ele17\_CaloIdT\_TrkIdVL\_CaloIsoVL\_TrkIsoVL\_Ele8\_CaloIdT\_TrkIdVL\_CaloIsoVL\_TrkIsoVL$
- For muon-electron events:
  - $HLT\_Mu8\_Ele17\_CaloIdT\_CaloIsoV\_L\_TrkIdV\_L\_TrkIsoVL$.

All of them require well reconstructed pairs of muons (or isolated electrons) where one of them has p_T>17 GeV (E_T>17 GeV for electrons) and the other one p_T>8 GeV (E_T>8 GeV).

The efficiency of the leptons to pass the double triggers are estimated in data with the tag&probe technique (as defined in section 7.1), in bins of (p_T,η), to take into account the different behaviour in different kinematical regions.

Trigger is not required in the MC, avoiding any dependence on the details of the trigger emulations. Instead, simulated events are weighted again with the efficiencies measured in data. On average, the dimuon trigger is 88% efficient whereas the dielectron trigger reaches the 94%.

Table 8.1 shows the efficiency of the double muon trigger, for the different bins of (p_T,η). Table 8.2 shows the efficiency for each of the two legs of the di-electron system. The offline lepton selection is tighter than the trigger requirements, as described in section 8.3, so it is unknown which electron corresponds to which trigger leg. So the efficiency of the double trigger is estimated as

$$\epsilon = \epsilon_{hlt17}(l_1) \cdot \epsilon_{hlt8}(l_2) + \epsilon_{hlt17}(l_2) \cdot \epsilon_{hlt8}(l_1) - \epsilon_{hlt17}(l_1) \cdot \epsilon_{hlt17}(l_2) \tag{8.2}$$

where ϵ_{hlt17} and ϵ_{hlt8} corresponds to the efficiencies for the *hlt17* and *hlt8* leg, respectively, assigned to the two electrons l_1, l_2. The last term eliminates double counting.

muon η	$0.0 < \|\eta\| < 0.9$	$0.9 < \|\eta\| < 1.2$	$1.2 < \|\eta\| < 2.1$	$2.1 < \|\eta\| < 2.4$
$0.0 < \|\eta\| < 0.9$	0.938 ± 0.011	0.880 ± 0.014	0.864 ± 0.012	0.880 ± 0.021
$0.9 < \|\eta\| < 1.2$	0.880 ± 0.014	0.836 ± 0.021	0.824 ± 0.017	0.819 ± 0.047
$1.2 < \|\eta\| < 2.1$	0.864 ± 0.012	0.824 ± 0.017	0.813 ± 0.010	0.804 ± 0.021
$2.1 < \|\eta\| < 2.4$	0.880 ± 0.021	0.819 ± 0.047	0.804 ± 0.021	0.784 ± 0.063

Table 8.1: Dimuon trigger efficiencies, for two tight muons, both with p_T>20 GeV, in four bins of pseudorapidity for each of the two muons.

η coverage	p_T range (GeV)	efficiency (HLT Ele8)	efficiency (HLT Ele17)
$0.0 < \|\eta\| < 0.8$	$20 < p_T < 40$	0.986 ± 0.001	0.983 ± 0.001
$0.8 < \|\eta\| < 1.4$		0.936 ± 0.001	0.932 ± 0.001
$1.6 < \|\eta\| < 2.0$		0.901 ± 0.002	0.895 ± 0.002
$2.0 < \|\eta\| < 2.5$		0.944 ± 0.002	0.933 ± 0.002
$0.0 < \|\eta\| < 0.8$	$40 < p_T < 200$	0.991 ± 0.001	0.989 ± 0.001
$0.8 < \|\eta\| < 1.4$		0.976 ± 0.001	0.972 ± 0.001
$1.6 < \|\eta\| < 2.0$		0.945 ± 0.002	0.938 ± 0.002
$2.0 < \|\eta\| < 2.5$		0.962 ± 0.002	0.951 ± 0.002

Table 8.2: Working point loose to the HLT Ele17 leg and HLT Ele8 tag&probe efficiencies in data.

8.3 Object selection

The selection of events done during the analysis is tighter than the trigger requirements, trying to select good Higgs candidates and to optimize the S/B ratio

to get the best possible sensitivity. A cut-based strategy on different variables is performed. A first step of cuts is called "preselection" and concerns kinematic cuts on the objects involved: electrons, muons and jets, independently. A final selection is done right after, based on combined information of the latter, like angular correlations of objects or invariant masses of pairs of objects.

8.3.1 Lepton selection

A pair of good isolated electrons or muons with opposite charge are required. Muons must be global muons satisfying the tight purity requirements[1]. A p_T threshold is set on p_T>40 GeV for the leading muon (the one with highest p_T) and on p_T>20 GeV for the subleading one. Forward muons with a pseudorapidity of $|\eta|$ >2.4, outside the tracker acceptance, are excluded. A PF relative isolation[2] is demanded, with a cut of Iso_{PF} <0.12 of energy measured in a cone or R=0.4 around the muon. It is corrected for pile-up dependences with the $\Delta\beta$ corrections. Figure 8.4 shows the p_T distribution for both muons after "preselection", comparing data and the estimated background.

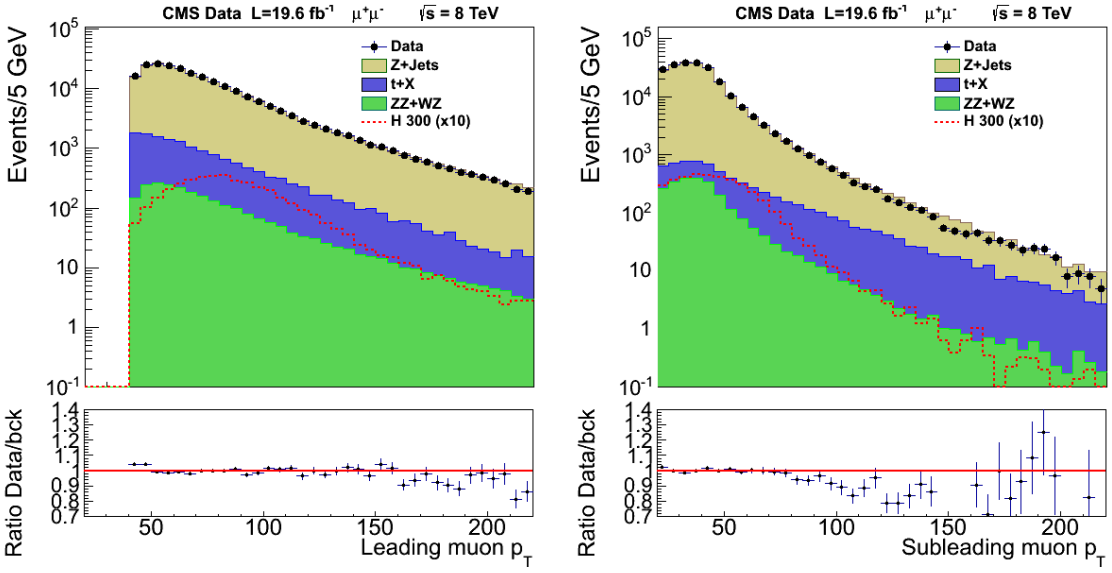


Figure 8.4: p_T distribution for the leading (left) and subleading (right) muon, in log-scale. Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background.

Electrons satisfy the Loose Working Point conditions[3], in order to maximize the efficiency. However, additional cuts are required to match some tighter criteria present

[1]These requirements are described in section 7.1.

[2]The muon isolation techniques and its corrections for pile-up dependencies are described in section 7.1.1.

[3]Detailed in section 7.2.

in the trigger. These cuts are relative to a detector-based isolation, which is the one implemented in the trigger, and are summarized in table 8.3.

Cut	Barrel	Endcaps
$\text{Iso}_{ECAL}/p_T<$	0.2	0.2
$\text{Iso}_{HCAL}/p_T<$	0.2	0.2
$\text{Iso}_{TRACKER}/p_T<$	0.2	0.2

Table 8.3: Additional cuts on electrons over the Loose Working Point to match some tighter criteria present in the trigger.

The same p_T>40/20 GeV is required for the leading/subleading electron, and a pseudorapidity $|\eta|$ <2.5. A gap region in the barrel-endcap ECAL overlap 1.4442> $|\eta_{SC}|$ >1.566 is excluded, where η_{SC} is the pseudorapidity of the ECAL supercluster which absorbed the electromagnetic cascade. A PF relative isolation, with a cut of Iso_{PF} <0.15 for a cone R=0.3, is corrected with the Effective Areas technique to avoid dependence on pile-up. Figure 8.5 shows the p_T distribution for both electrons after preselection.

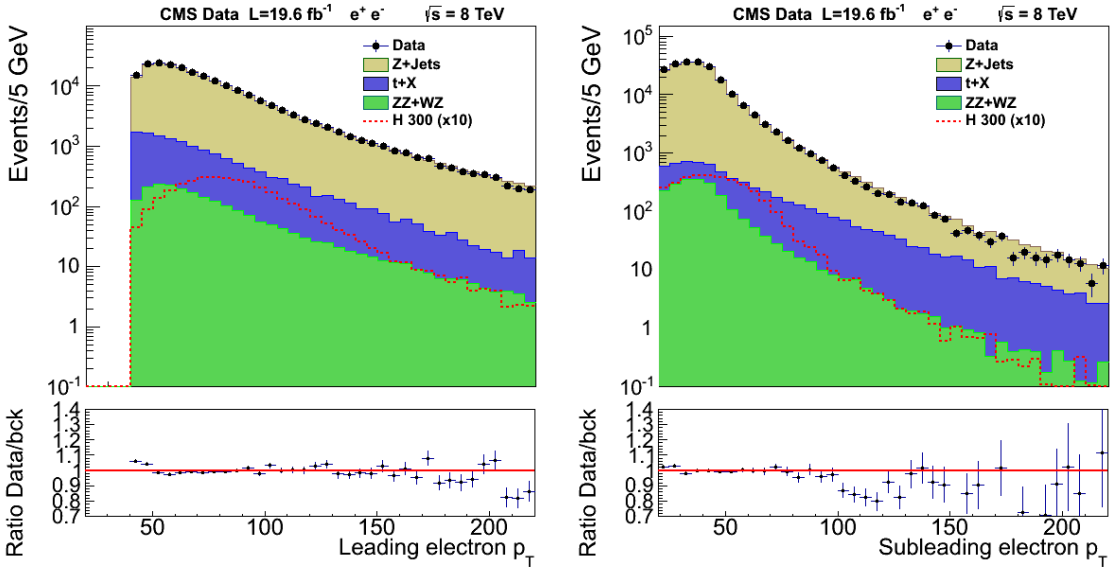


Figure 8.5: p_T distribution for the leading (left) and subleading (right) electron, in log-scale. Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background.

The two leptons must derive from a Z decay. Hence, the invariant mass of the system l^+l^- is reconstructed, to select only events with the mass around the Z nominal mass: 76< m_{ll} <106 GeV. This cut reduces the background yields ∼15%.

Figure 8.6 shows the m_{ll} distribution after the selection, for $Z \to e^-e^+$ (left plot) and $Z \to \mu^-\mu^+$ (right plot). Small differences between data and background simulation

are due to imperfect lepton energy scale and resolution calibration of the MC. However, this does not affect the final analysis at all. The m_{llqq} resolution is dominated by the jet resolution, which is far worse than the lepton resolution and the observed differences.

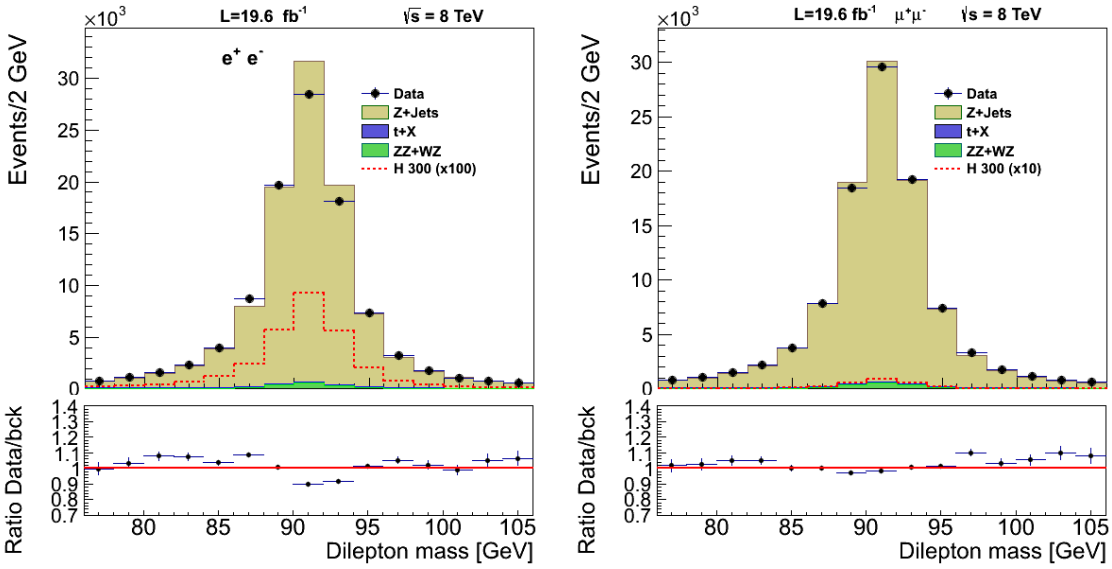


Figure 8.6: Invariant mass of the dilepton system for e^+e^- (left) and $\mu^+\mu^-$ (right). Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background.

8.3.2 Jet Selection

Particle-flow jets built with the anti-k$_T$ algorithm and a radius R=0.5 are selected, including the L1+L2+L3 corrections, as described in section 7.3. p_T and η thresholds are set on p_T>30 GeV, and $|\eta|$ <2.4, respectively, thus allowing high reconstruction efficiency and precise energy measurements using PF techniques. An angular separation of $\Delta R=\sqrt{(\Delta\phi)^2+(\Delta\theta)^2}$ ≥0.5 among any jet with any of the leptons is required, to discard the possibility of any of the leptons to be misidentified also as a jet, and consequently being double counted as lepton and jet in the same Higgs candidate.

An additional loose jet identification criteria is required, based on the energy fractions of the different jet components (neutral hadrons, photons, electrons, charged hadrons) and the number of particles inside the jet (Table 8.4), to assure good quality of jets.

One of the consequences of the increase of pile-up is that the number of jets also increases, and the quality of the jets worsens, as particles from other collisions are more likely to be included in the jet-reconstruction algorithm. It is impossible to guarantee if a jet is produced by a Z decay or not, but it is possible to reduce the pileup contamination and the "bad-reconstructed" jets. The β variable exploits the

$E_{neutral\ had.}$	<100%
E_{photon}	<100%
$E_{charged\ had}$	>0.0%
$E_{electron}$	<100%
#charged particles	>0
#charged+neutral particles	>1

Table 8.4: Jet identification criteria to improve the quality of jets.

particle-flow algorithm to calculate the fraction of charged particles coming from the primary vertex in a jet:

$$\beta = \frac{\#\ tracks\ inside\ the\ jet\ coming\ from\ the\ primary\ vertex}{\#\ total\ tracks\ inside\ the\ jet} \tag{8.3}$$

A jet built with high contribution from pile-up radiation will show values of β close to zero, whereas a hard jet coming from the primary vertex, will show values close to one.

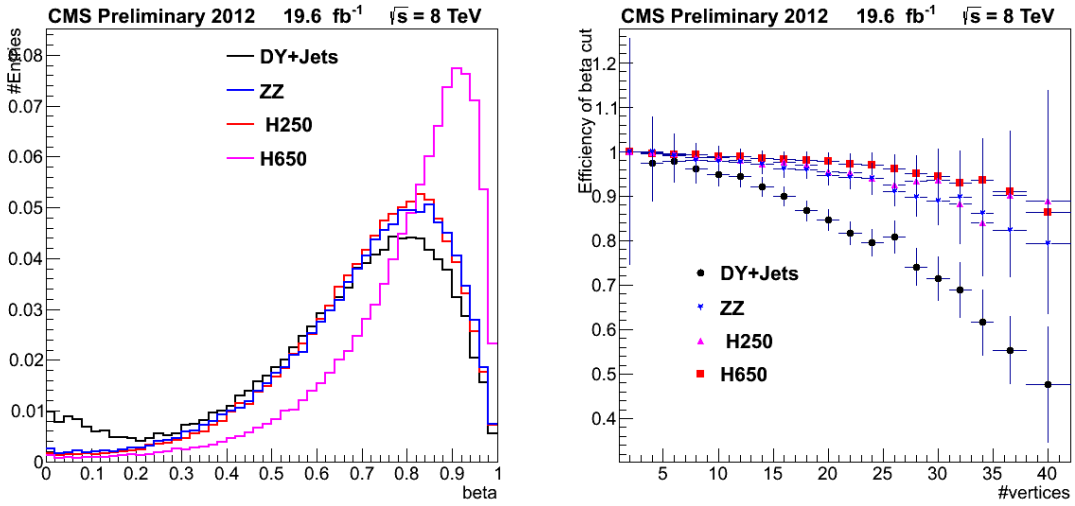


Figure 8.7: Left:β variable for different processes in the *llqq* final state. Right: Efficiency of a β >0.2 requirement on both jets versus the number of reconstructed vertices, for different processes, in the *llqq* final state.

Left plot in Fig. 8.7 shows the distribution of the β variable for two signal and two background processes. Jets from a Z decay have values closer to 1, as they are coming from the primary vertex and the jet is well defined. The more massive the Higgs hypothesis, the harder the jets will be, so $\beta \to 1$, as observed comparing the distribution for $m_H = 250$ and $m_H = 650$. $t\bar{t}$ decays do not suffer from significant pile-up contamination either, as the jets are produced as part of the basic production process.

On the other hand, Z+jets jets may be softer and worse quality. Its distribution has a minimum around β ∼0.2 and a sizable fraction of events close to zero. These jets are reconstructed from pile-up components and contaminate the light flavour categories of the analysis.

Right plot in Fig. 8.7 shows the efficiency of a β >0.2 cut on both jets for the different processes. While the signal is almost not affected, a reduction of 20% of the Z+Jets events is gained from 20 reconstructed primary vertices, (remember the average pile-up during 2012 was 21). This is very important, as Z+jets is, by far, the main background. Consequently, a requirement of $\beta > 0.2$ is demanded in both jets.

Figure 8.8 shows the p_T distribution for both jets after preselection, for data and background estimations.

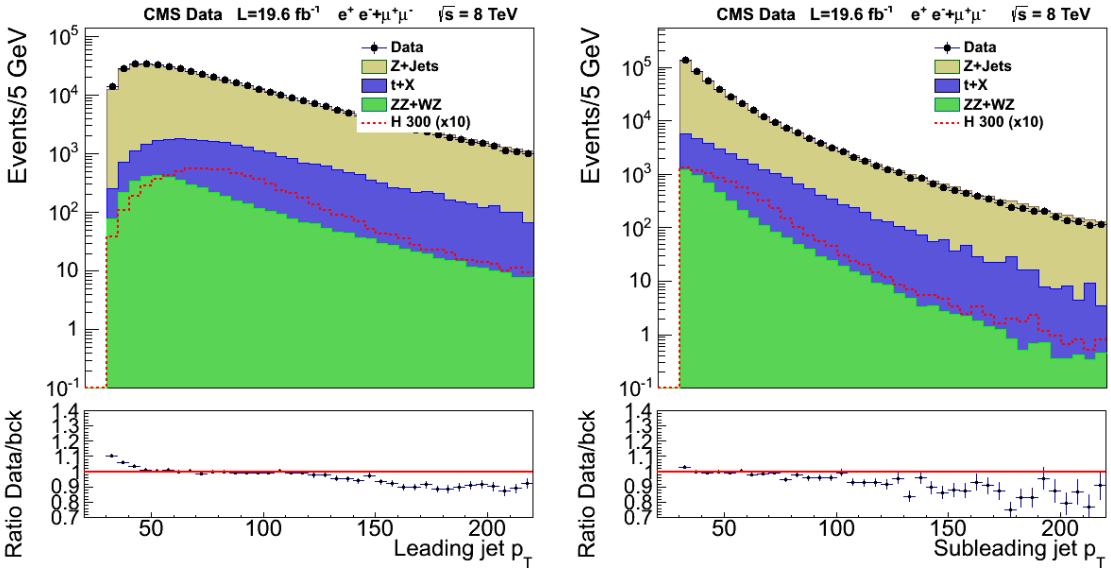


Figure 8.8: p_T distribution for the leading (left) and subleading (right) jets, in log-scale. Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background.

Figure 8.9 shows the m_{jj} distribution after selection cuts. The resolution of m_{jj} is worse than the leptonic m_{ll} and this determines the resolution of the m_{llqq} spectrum. The two jets must be also resonant with the Z mass. Hence, the signal window is defined in the m_{jj} spectrum 71<m_{jj}<111 GeV. Only ∼30% of the events satisfying all the selection criteria belong to the signal region. A Control region is defined at both sides of the peak, and will be used in the Z+jets background determination, described in section 8.4 The control region (also referred as "sidebands" region) covers 60<m_{jj}<71∪111 <m_{jj}<130 GeV.

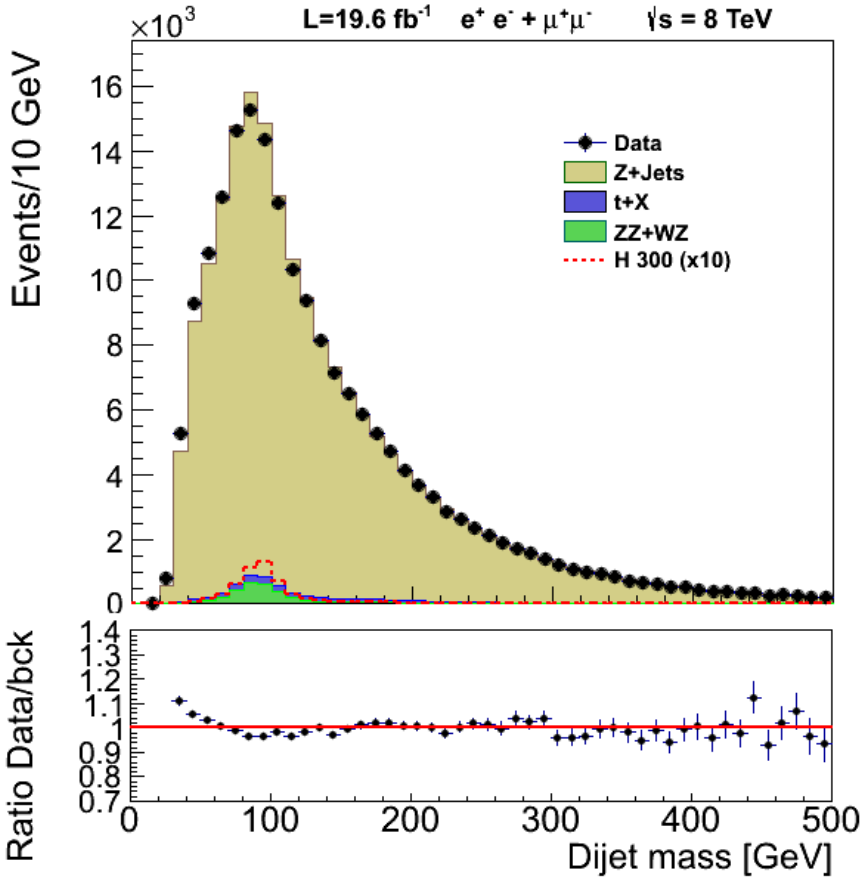


Figure 8.9: Distribution the dijet invariant mass m_{jj} after selection cuts. Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background. The signal region is defined in the range m_{jj} spectrum $71<m_{jj}<111$ GeV and the control region is defined in the sidebands $60<m_{jj}<71 \cup 111<m_{jj}<130$ GeV.

8.3.3 Angular discriminant

After the preselection of muons, electrons and jets satisfying our quality criteria, a Higgs candidate is built from the quadrimomentum of the four objects. In case that more than two jets or leptons pass the preselection, more than one Higgs candidate is obtained per event, according to the possible combination of two leptons + two jets.

There exist different features in the Higgs decay that allow to discriminate signal from background events. The fact that the Higgs boson is a particle with spin 0, decaying into two Z bosons with spin 1, decaying in turn to four fermions, constrains the kinematics in the decay. It has been shown [93, 94] that the decay chain $ab \to X \to ZZ \to 2l2q$ is fully described by five correlated angles in the decay, shown in Fig. 8.10. The two axes $\hat{z}$ and $\hat{z'}$ refer to the parton collision and H → ZZ decay axis in the Higgs rest frame, respectively. θ^* is the angle between both axes. θ_1 and

θ_2 are the angles between the l^- (q) from the $Z_1 \to ll$ ($Z_2 \to qq$) and the opposite direction of the H, in their Z rest frame. ϕ is the angle between the production and decay planes of the two Z systems. Finally, ϕ_1 is the angle between the production plane and the Z_1 decay plane. The five angle distributions are shown in Fig. 8.11. In background events, where jets come from QCD processes, the angles are completely uncorrelated, which allows to separate signal from background.

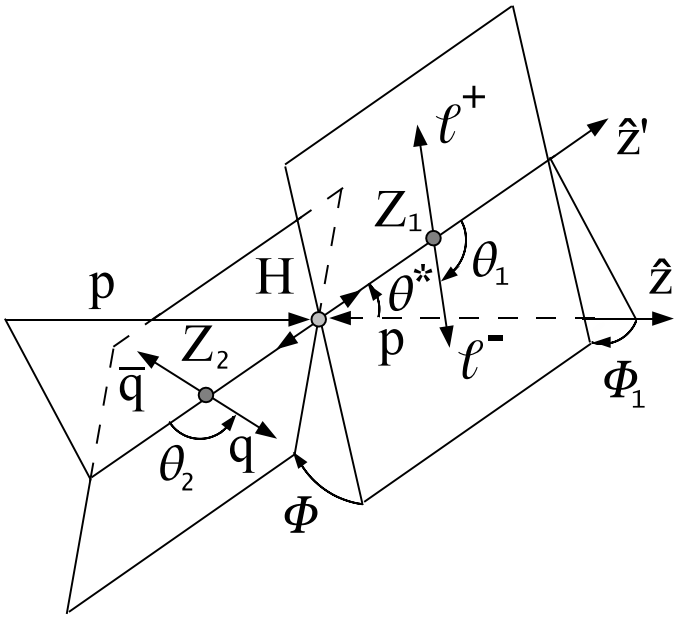


Figure 8.10: Diagram representing the $H \to ZZ \to l^+l^-q\bar{q}$ decay and the five angles describing it.

Angular variables are weakly correlated with the Z and Higgs masses, contrary to other kinematical variables, so the shape of the m_{ZZ} distribution, used as final discriminant, is not biased.

A linear likelihood discriminant, LD, is constructed from the signal and background probabilities defined using the five angles, as in Eq. 8.4:

$$LD = \frac{\mathcal{P}_{sig}}{\mathcal{P}_{sig} + P_{bck}} \tag{8.4}$$

where $\mathcal{P}_{sig}$ and $\mathcal{P}_{bck}$ are the probability density functions (p.d.f) of the five helicity angles for signal and background respectively.

For the signal, the probabilities are considered fully correlated and corrected by an acceptance function for each of the angles, estimated from simulation.

$$\begin{aligned}\mathcal{P}_{sig} = \mathcal{P}\left(\theta_1, \theta_2, \theta^*, \phi, \phi_1 : m_{ZZ}\right) \times \mathcal{G}_{\theta_1}\left(\theta_1 : m_{ZZ}\right) \times \mathcal{G}_{\theta_2}\left(\theta_2 : m_{ZZ}\right) \times \\ \times \mathcal{G}_{\theta^*}\left(\theta^* : m_{ZZ}\right) \times \mathcal{G}_{\phi}\left(\phi : m_{ZZ}\right) \times \mathcal{G}_{\phi_1}\left(\phi_1 : m_{ZZ}\right)\end{aligned} \tag{8.5}$$

Probabilities for background are estimated from simulation and considered fully uncorrelated, as stated in Eq. 8.6:

$$\mathcal{P}_{sig} = \mathcal{P}_{\theta_1}(\theta_1 : m_{ZZ}) \times \mathcal{P}_{\theta_2}(\theta_2 : m_{ZZ}) \times \mathcal{P}_{\theta^*}(\theta^* : m_{ZZ}) \times \mathcal{P}_{\phi}(\phi : m_{ZZ}) \times \mathcal{P}_{\phi_1}(\phi_1 : m_{ZZ}) \tag{8.6}$$

Signal and background p.d.f are parametrized as a function of m_{ZZ}. Plot at the bottom-right of Fig. 8.11 shows the distribution for data and background estimation after the final selection.

An optimized threshold of LD$\geq$0.5 is established for all categories, which reduces the Z+jets background by a factor of two, while retaining a signal efficiency of at least 80% for all m_H.

8.3.4 Missing transverse energy

The $t\bar{t}$ background is especially important in the H $\rightarrow$ ZZ $\rightarrow$ llb$\bar{\text{b}}$ final state, when the two W$^{\pm}$ decay leptonically. However, the presence of two neutrinos in the final state facilitates the discrimination. The Higgs decays are well balanced in transverse momentum, so no $\not\!\!E_{\text{T}}$ is expected. The $\not\!\!E_{\text{T}}$ significance defined in Eq. 7.6 is selected to reduce the $t\bar{t}$ contamination, and shown in Fig. 8.12. A loose cut λ <10 is applied to the events, which has almost no effect in the categories with no b-tagged jets, but implies a $\sim$20% reduction of the background in the 2 b-tag category. Small discrepancies in the data-MC comparison arise in the very low $\not\!\!E_{\text{T}}$ region ($\lambda < 6$). A 15% corrector factor is applied to those MC where no real $\not\!\!E_{\text{T}}$ is expected (Z+jets, dibosons, Higgs production). This correction implies less than 1% of efficiency loss and has been validated in regions with large values of $\not\!\!E_{\text{T}}$.

Table 8.5 summarizes all the selection criteria performed on the recorded dataset. The resulting yields for the observed data and expected background and signal are depicted in table 8.6.

8.3.5 b-tagging identification and final selection

Jets in the Z+jets background are preferentially originated by light flavour quarks and gluons, result of the large gluon radiation, and the predominance of the u and d valence quarks inside the proton. In comparison, the Z decay includes a higher fraction of heavy flavour jets. This is exploited by splitting the sample according to the number of b-tagged jets in the event: 0, 1 or 2 b-jets. The 2-btag category has much lower statistics, but with a higher S/B, whereas the 0-btag category includes most of the sample, but with a poorer S/B. The combined sensitivity of the three exclusive channels (actually six, considering e^+e^- and $\mu^+\mu^-$ separately) is therefore improved.

The Jet Probability (JP) algorithm, mentioned in section 7.3.1, is used. Figure 8.13 shows the distribution of the tagger discriminating variable. Two different working points are considered: for the 1-btag category, a tagger value d>0.275 is required,

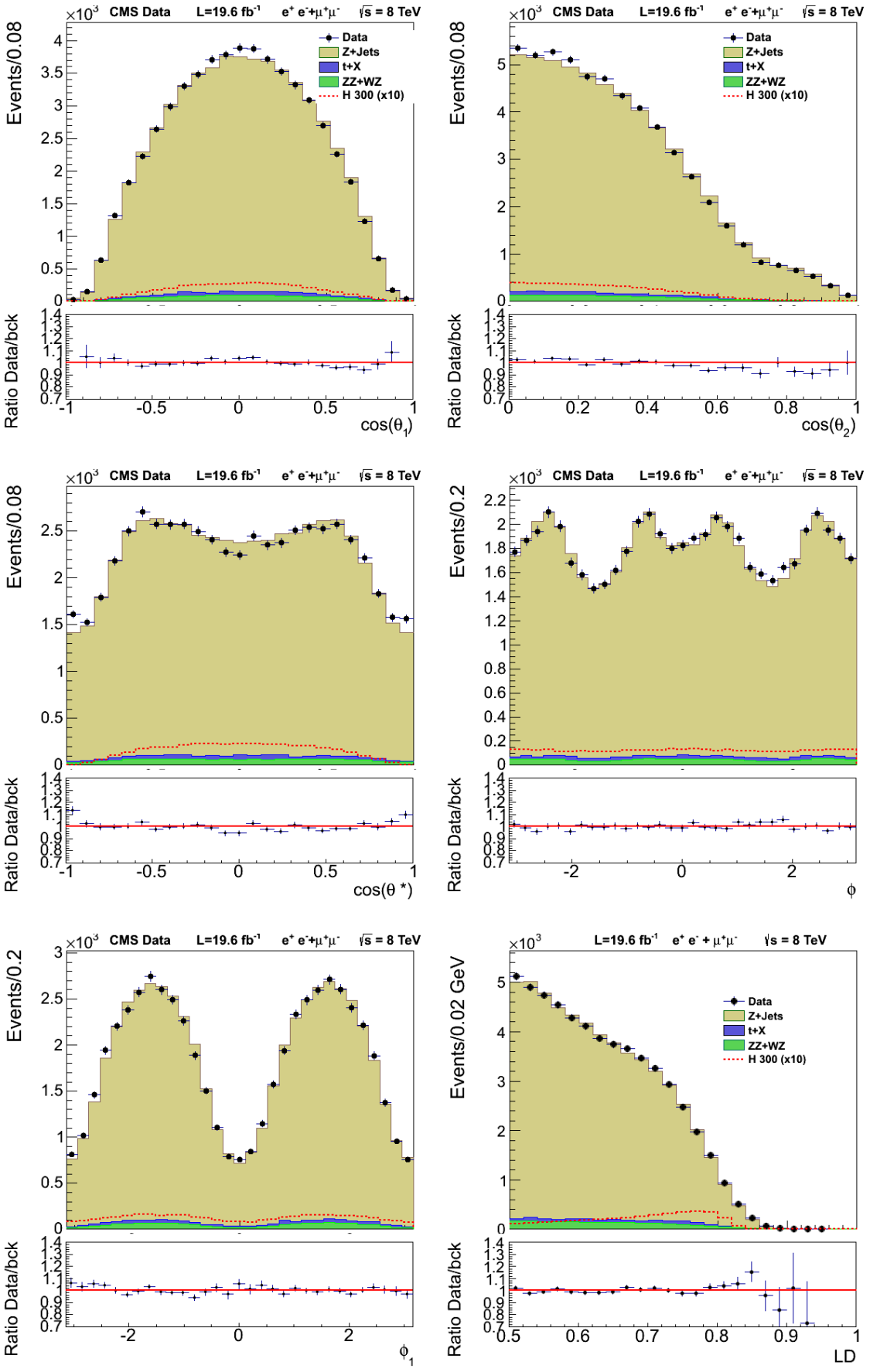


Figure 8.11: 5 angular distributions of $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, ϕ, ϕ_1 and the angular likelihood discriminant, for final selection except for the LD discriminant cut.

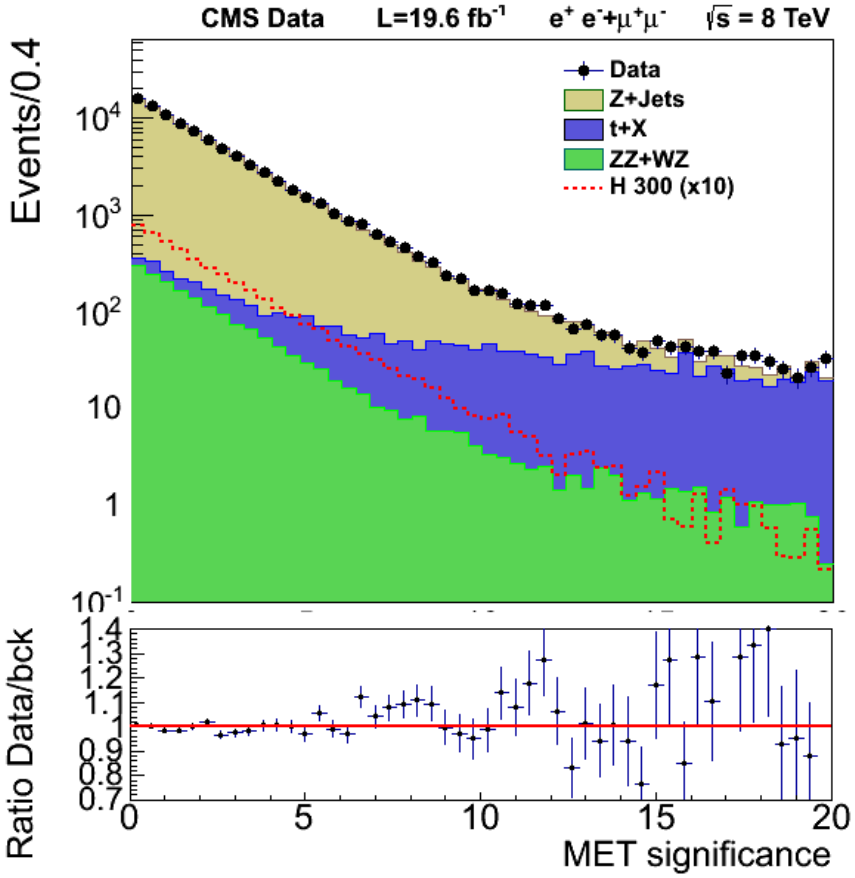


Figure 8.12: Particle-flow missing transverse energy significance in data and simulation for events with electrons and muon combined, after the full selection except the cut on MET significance. Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background.

corresponding to the Loose Working Point (JPL). It corresponds to a 80% efficiency of b-tagging (see Fig. 7.8). For the 2 b-tag category, one Loose b-tag and one Medium btag (JPM) are demanded. The Medium Working Point corresponds to d>0.545 and 65% efficiency.

Simulation does not perfectly reproduce the behaviour of the algorithm in data. Scale factors data/MC are computed in bins of (p_T,η) for the b-tag efficiency, as well as for the misidentification of light jets as bjets (mistags). The tagging is used to categorize events, not to filter them, so instead of applying a weight to each event, migrations from one category to another are allowed. Each simulated jet is first labelled as light, loose b-jet, or medium b-jet as outcome of the JP algorithm. Then, efficiencies and mistag efficiencies are compared with generated pseudorandom numbers allowing the jet to migrate to another level, more or less pure.

Only one Higgs candidate is allowed in each event. Candidates in the signal region

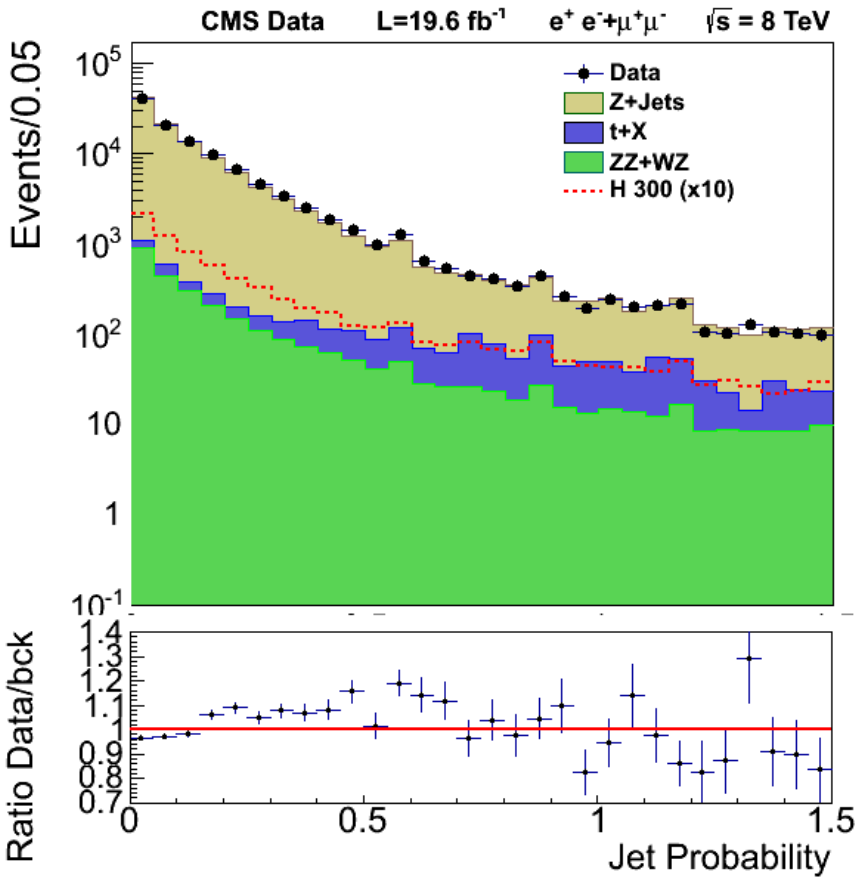


Figure 8.13: Discriminant of the Jet Probability algorithm for b-tagging after all selection cuts.

are picked in the first place. If more than one candidate in the signal region has passed all the selection criteria, which happens about the 3% of the times, the candidate in the category with more b-tagged jets is chosen. If more than one candidate still remains, the one with the m_{jj} and m_{ll} closer to the Z nominal mass will be picked, by minimizing the quantity: $|m_{jj} - M_Z| + |m_{ll} - M_Z|$.

In case that no final candidate is found in the signal region, a candidate in the sidebands is selected, following the same preference criteria.

Figure 8.14 shows the final yields for the different b-tag categories.

8.4 Background determination

This section refers to the estimation of the three background processes considered after the final selection: Z+jets, $t\bar{t}$ and diboson production (ZZ,WZ,WW). Other contributions like W+jets or multijet events where jets are misidentified as leptons are negligible due to the tight dilepton selection. The overwhelmingly dominant

observable	0 btag	1 btag	2 btag
Preselection			
muon quality	global tight muon		
electron quality	loose electron		
jet quality	taggable		
same lepton flavour, opposite charge			
$p_T(l^{\pm})$	$> 40/20\ GeV$		
p_T(jets)	$> 30\ GeV$		
$\|\eta\|(\ell^{\pm})$	$(e^{\pm}) < 2.5$, $(\mu^{\pm}) < 2.4$		
$\|\eta\|$(jets)	< 2.4		
jet β	> 0.2		
ΔR	> 0.5		
Final selection			
b-tag	none	JPL	JPM & JPL
$m_{\ell\ell}$	[76,106] GeV		
m_{jj}	[71,111] GeV		
helicity LD	> 0.5		
missing $E_T$significance	< 10		

Table 8.5: Summary of requisites in the event selection.

	0 btag		1 btag		2 btag	
	$\mu\mu jj$	$eejj$	$\mu\mu jj$	$eejj$	$\mu\mu jj$	$eejj$
expected background	14857	13930	5490	4929	526	466
observed data	14745	13787	5470	4974	523	480
m_H (GeV)	signal expectation					
250	114.1	107.3	54.5	51.8	18.3	17.6
300	127.5	119.0	64.7	56.7	24.6	21.6
400	125.5	114.6	64.9	58.8	27.3	24.9
500	59.5	55.8	31.0	28.9	13.7	12.7
600	22.6	21.3	12.3	11.5	5.4	5.0

Table 8.6: Expected and observed yields with 19.6 fb^{-1} of data. The expected background is composed of p_T-weighted simulated Z+jets, data-driven top+X and diboson MC. The Signal yields include gg and VBF production modes. The background is normalized to the data in the m_{jj} sideband.

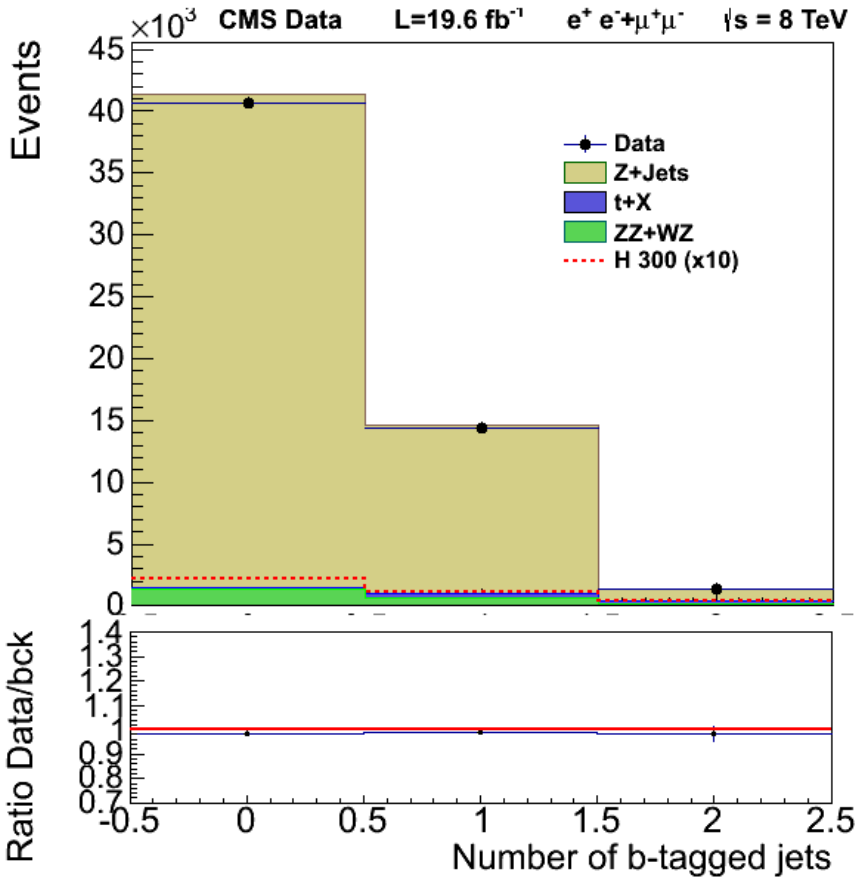


Figure 8.14: Number of events in each btag category after full selection, for the electron and muon channel combined, in linear scale. Dots indicate data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue top background.

background is the Z+jets, even though jets do not proceed from a Z.

The Z+Jets distribution shape is taken from the simulated Z+(n) parton samples, detailed in appendix A. The corresponding m_{llqq} distribution is essentially a falling exponential above 220 GeV. Below that value, acceptance effects give rise to a steep edge difficult to reproduce with simulations, therefore reducing the sensitivity of the search.

MC simulation reproduce very well the data for all observables considered in the control region. Only small differences arise in the p_T distribution of the lljj system for low p_T values, attributed to a bad modelling of the p_T spectrum. This is shown in the left-top plot of the Fig. 8.15, where the ratio data/MC is represented for both the signal region and the sideband region. Apparent discrepancies in the tail of high p_T values are simply statistical fluctuations due to the decreasing number of candidates with the p_T.

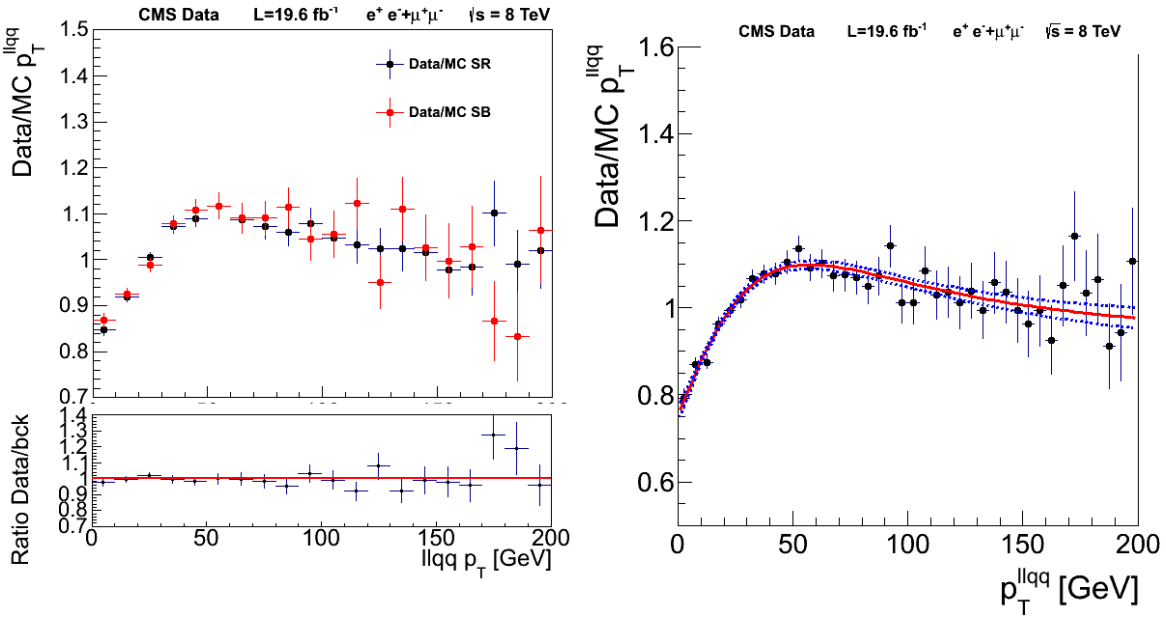


Figure 8.15: Left-top: data over Z+jets ratio of the p_T distributions of the $lljj$ system in the m_{jj} sideband (red triangles) and signal (black dots) regions. Left-bottom: ratio of the ratios above. Right: ratio of the data over simulation p_T distributions together with the fit to $f(x)$ (red solid line). The blue dashed lines are the $\pm 1\sigma$ statistical error bands calculated propagating the full correlation matrix of the fit.

This behaviour in MC is similar in the sidebands and in the signal region, compared to the data in the sidebands, as depicted in the left-bottom plot of Fig. 8.15). It shows the ratio of the ratio distributions in the left-top plot of the same figure. Consequently, a correction is applied to the MC from the ratio data/MC in the sidebands.

The p_T spectrum is therefore parametrized to the function

$$f(p_T) = \left(1 + \frac{1}{a + bp_T^2}\right) \frac{1}{e^{-p_T} + 1} \tag{8.7}$$

in data and MC and a p_T-dependent weight is applied to the simulated events. The fit to the ratio is shown in the right plot of Fig. 8.15. $t\bar{t}$ and diboson contributions to the data in the sidebands was substracted from their respective estimations, to get an unbiased correction of the Z+jets. The same correction is applied to all b-tag categories, after it was checked the consistency with category−dependent corrections within errors.

In order to be independent of the b-fraction simulation of the Z+Jets MC, the correct normalization of the MC for the different categories in the signal region is constraint to the relative normalization data-MC in the m_{jj} sidebands (SB), after substracting the $t\bar{t}$ and diboson contribution in the SB. It is performed independently for the three categories and for each lepton flavour.

The $t\bar{t}$ background is estimated on a data-driven basis. This avoids relying on the emulation of the process and avoids dealing with several associated uncertainties.

The two leptons produced in the $t\bar{t}$ decay are not originated from the same particle, so the final state with lepton flavour asymmetry $e^{\pm}\mu^{\mp}q^+q^-$ is a good control region to estimate the top background in the signal region: it is signal free; its kinematics is expected to be identical to the $t\bar{t}$ contribution in the signal region; the $e^{\pm}\mu^{\mp}$ normalization is expected to be the same as the $e^+e^- + \mu^+\mu^-$; and moreover it accounts for other small backgrounds (as WW+jets, $Z \to \tau^+\tau^-$ + jets, single top, fakes) where the lepton flavour symmetry can be invoked as well.

The method has been validated with the Top Powheg+Pythia simulation and in a data control region with large top contribution. This region is defined in the m_{ll} spectrum, out of the Z invariant mass $|m_{ll} - M_Z| > 20$ GeV, to remove signal, dibosons and Z+jets background. Large $\not{E}_T$ significance $\lambda > 8$ is also required.

$e^{\pm}\mu^{\mp}$ shape has been validated using the top simulation. Left plot in Fig. 8.16 compares the m_{llqq} distribution for $e^{\pm}\mu^{\mp}$ data to the $t\bar{t}$ MC. The agreement in normalization $e^{\pm}\mu^{\mp}$ vs $e^+e^- + \mu^+\mu^-$ is better than 1±0.01% in MC. In data control region the The $e^{\pm}\mu^{\mp}$ vs $e^+e^- + \mu^+\mu^-$ normalization is validated in the top data control region, as depicted in the right plot of Fig. 8.16. The number of events of the $e^{\pm}\mu^{\mp}$ data equals the $e^+e^- + \mu^+\mu^-$ top subsample within statistical errors.

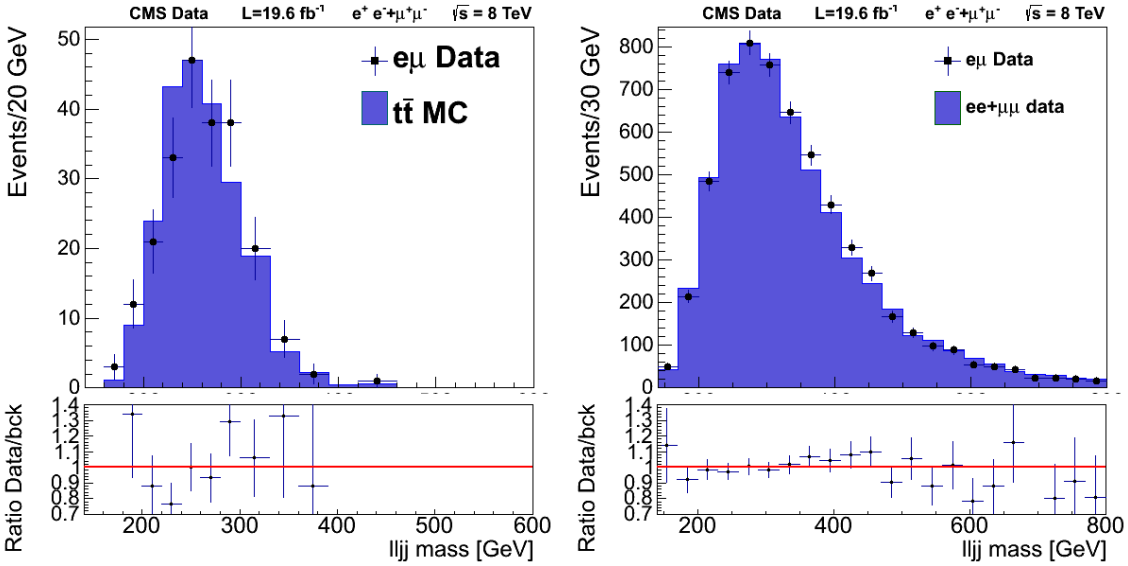


Figure 8.16: Left: m_{llqq} invariant mass for $e^{\pm}\mu^{\mp}$ data compared to the prediction of $t\bar{t}$MC plus other small backgrounds; red dots are $e^{\pm}\mu^{\mp}$ data; white histogram, $t\bar{t}$ MC. Right: m_{llqq} invariant mass for $e^+e^- + \mu^+\mu^-$ and $e^{\pm}\mu^{\mp}$ data for events outside the leptonic Z mass window, 2 b-tagged jets, and $\lambda > 8$; black dots are $e^+e^- + \mu^+\mu^-$ data; blue histogram, $e^{\pm}\mu^{\mp}$data.

The fake component is estimated from $e^{\pm}\mu^{\mp}$ data; the yield of events with one or two non-isolated leptons (in the combined relative isolation region I=[0.25, 0.85]), is extrapolated into the isolated lepton region assuming a flat distribution of the combined relative isolation variable.

Finally, diboson contribution (WZ,ZZ) is extracted directly from simulation but its effect is irrelevant due to the small cross section of these processes.

8.5 Signal determination

As explained in section 3.2.2, the Higgs boson width depends on its mass, and a narrow resonance scenario approximated by a Breit-Wigner function is no longer valid at higher masses (about m_H>400 GeV) [95]. A more correct approach to describe the Higgs invariant-mass distribution, known as Complex Pole Scheme (CPS), has been proposed [31, 92]. This more accurate description affects the cross section and also the signal lineshape.

At high Higgs mass, the interference between the Higgs signal and the gg→ ZZ background becomes large. The effect of interference has been shown to be constructive below the Higgs mass peak and destructive above. It has therefore a negligible effect on the total cross-section ($1 - 2\%$) but it modifies the ZZ invariant mass distribution [96]. The lineshape is also corrected from this feature.

The present analysis has been optimized for the dominant $gg \rightarrow H$ production mode. However, the vector boson fusion (VBF) production represents about 10% of the total cross-section, so it is also included in the analysis, as it would exist in the data. A dedicated analysis to study the VBF production of the Higgs boson in the $lljj$ final state is currently ongoing to take the specific features into account in the optimization of the search. Signal expectation is fully extracted from the simulation. The efficiency of the final selection described in section 8.3 is depicted in Fig. 8.17 for the six channels considered, as a function of m_H. Efficiency is defined as the number of events passing all the selection in one of the channels divided by the total number of events generated for the six channels. It is observed how the efficiency drops for $m_H > \sim$ 500 GeV. This is mainly consequence of the jet merging. A Higgs boson with a higher mass produces more boosted Z bosons. This implies that the angular separation between the two jets decreases, to the point that they may be reconstructed as only one jet. Consequently, efficiency diminishes as no 4-object can be reconstructed, or if any other third jet existing in the event is selected, it is less likely to pass the selection. Signal selection includes candidates whose jets are not those matching the jets generated from the Higgs decay.

The decrease in signal efficiency determines the upper range of m_H presented in this work, up to m_H=600 GeV. A dedicated analysis studying the case of two merged jets in the reconstruction is being developed in order to extend the range of this search up to 1 TeV.

8.6 Systematic uncertainties

Systematic uncertainties affecting the analysis have been evaluated. They concern both the normalization and shape of the signal and background distribution. Table 8.7

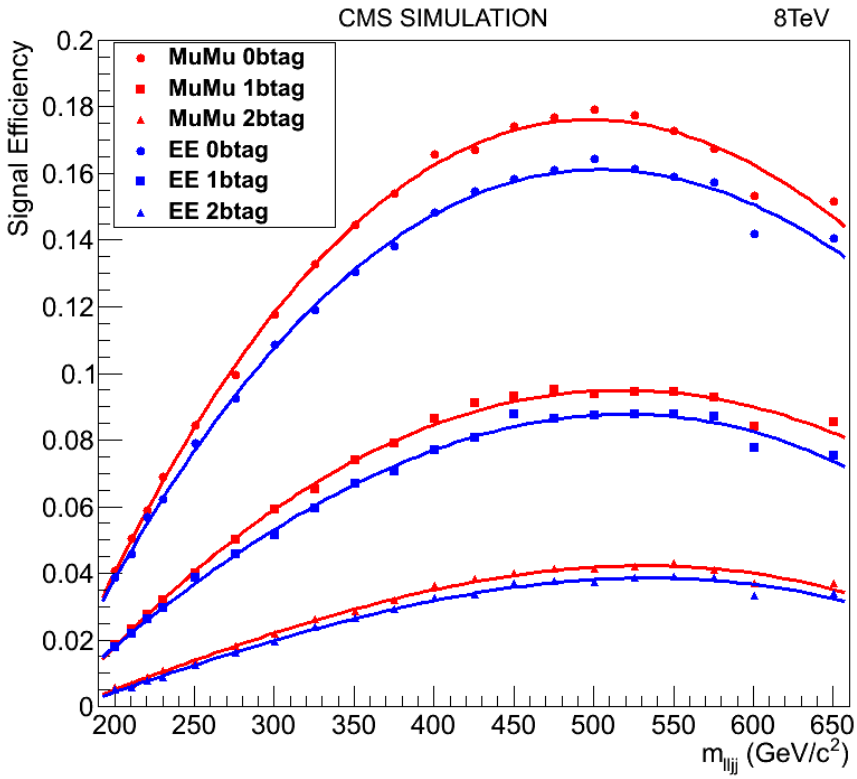


Figure 8.17: Parameterization of the signal efficiencies, as function of the Higgs mass hypothesis, in the three btag categories, in the muon and electron channels. The efficiency in a particular channel is defined as the yields in that channel after the final selection, divided by the total generated events in all the six channels.

summarizes the systematic uncertainties affecting the signal normalization, whereas the table 8.8 summarizes the systematic uncertainties related to the background estimation. This section briefly describes the different uncertainties and the evaluation method.

8.6.1 Luminosity

The uncertainty associated to the integrated luminosity measured by CMS has been evaluated in [97] by the Luminosity Working Group, yielding a value of 4.4%.

8.6.2 Pile-up

The MC is reweighted to match the pile-up distribution in data, and the number of pile-up measured in data is a source of systematic uncertainty. Both jet reconstruction and lepton isolation criteria are corrected from pile-up contamination on a event-by-event basis. The uncertainty, according to the CMS Pileup Working Group is measured

re-estimating the number of interactions after varying a ±5% the inelastic proton-proton cross section σ_{pp} =69.3 mb. Figure 8.18 shows the pile-up distribution for the central σ value and the ±5% variations. The MC distribution is again matched to the data distributions and the yields are recomputed in each case. The uncertainty obtained about 1%-2% on the signal normalization, whereas it has a very small effect on the background. Table 8.9 details the measured uncertainty for each category and for different m_H hypothesis.

8.6.3 Lepton identification, trigger, isolation, energy scale and resolution

Lepton trigger efficiencies (detailed in section 8.2.3) have been computed using the tag&probe technique, as well as identification and isolation efficiencies, as described in sections 7.2 and 7.2. For muons, the total normalization uncertainty is 2.7%, combining contributions from the trigger (2.5%), identification (1.0%), and isolation (0.4%). For electrons the total normalization uncertainty is 2%, dominated by identification (2%) and a much smaller contribution from the dielectron trigger efficiency.

Uncertainties related to the muon momentum scale, or electron energy scale are tiny, much below than 1% .

Source	0 b-tag	1 b-tag	2 b-tag	Comment
Muon trigger & ID	2.7%			Tag&probe study
Electron trigger & ID	2%			Tag&probe study
Electron energy scale	0.2%			
Muon momentum scale	0.1%			
Jet reconstruction	1-4%			JES, correlated among categories
b-tagging eff. and mistag rate	1-4%	1-5%	5-8%	Anti-correlated among categories
MET	$< 1\%$			Loose requirement
Pile-up	1-2%			Correlated between categories
Production mechanism (PDF)	1.5%			PDF4LHC, acceptance only
Production mechanism (lineshape)	0-3%			Only for $M_H > 400\ GeV$
Luminosity	4.4%			Same for all analyses
Higgs cross-section (for R)	13-15%			Detailed table from YR available

Table 8.7: Summary of systematic uncertainties on signal normalization. Most sources are multiplicative errors on the cross-section measurement, except for expected Higgs cross-section (which is relevant for the measurement of the ratio to SM expectation R).

Source	Normalization	Shape
Muon trigger & ID	2.7%	
Muon momentum scale	0.1%	
Electron trigger & ID	2.0%	
Electron energy scale	0.5%	
Jet energy scale	5.5%	0-4%
b-tagging efficiency SF 0-tag	+0.4%	
b-tagging efficiency SF 1-tag	-0.8%	
b-tagging efficiency SF 2-tag	-4.5%	
Mistag SF 0-tag	-1.9%	
Mistag SF 1-tag	+7.8%	
Mistag SF 2-tag	+6.2%	
MET	0.3%	
Pile-up	0.1%	
${p_T}^{\ell\ell jj}$ weighting	0.8%	0-3%
Diboson cross section	15%	
Luminosity	4.4%	
Residual difference		0-15% (0-btag)
data-background in		0-30% (1-btag)
control region		0-40% (2-btag)

Table 8.8: Summary of systematic uncertainties on the normalization and shape of the background determination.

Electron			
m_H [GeV/c^2]	$N_{\text{btag}} = 0$	$N_{\text{btag}} = 1$	$N_{\text{btag}} = 2$
200	$^{+1.57}_{-1.91}$	$^{+0.34}_{-0.14}$	$^{+0.01}_{+0.08}$
300	$^{+1.04}_{-0.93}$	$^{+0.99}_{-0.77}$	$^{+1.07}_{-1.01}$
400	$^{+0.38}_{-0.37}$	$^{+0.26}_{-0.31}$	$^{+1.64}_{-1.50}$
525	$^{+0.13}_{-0.15}$	$^{+1.07}_{-1.06}$	$^{-0.11}_{-0.06}$
600	$^{+1.03}_{-0.99}$	$^{+0.95}_{-0.96}$	$^{+0.16}_{-0.27}$

Table 8.9: The relative systematic uncertainty in percentage (%) from PU reweighing.

8.6.4 Jet energy scale and resolution

The most relevant uncertainty associated to the jet reconstruction is the jet energy scale (JES). The variation on the final efficiency is evaluated by recalculating the jet quadrimomentum after shifting the jet energy scale $\pm 1\sigma$ up and down of the measured JES uncertainty.

This has a direct effect in the jet p_T distribution, and therefore in the final acceptance, which depends on the m_{jj} invariant mass distribution.

The jet energy resolution had also been evaluated by smearing the p_T of the jets,

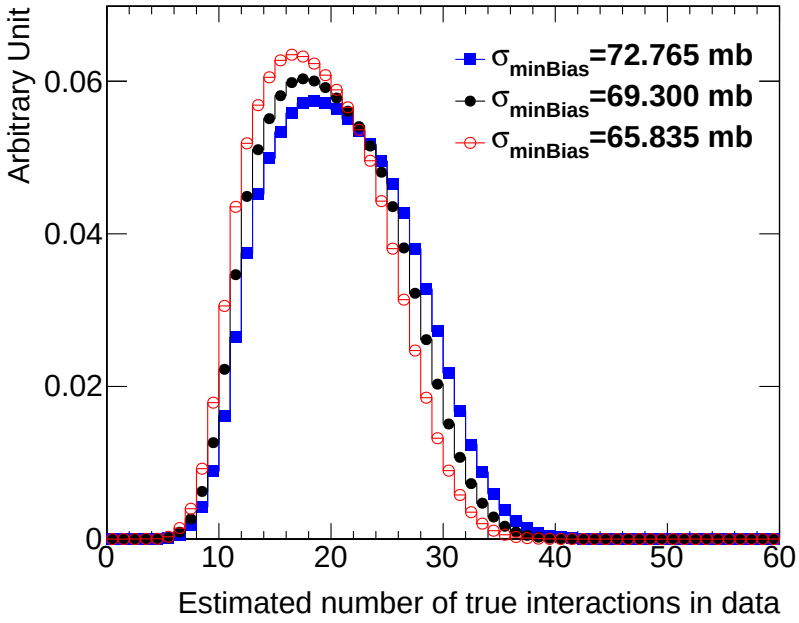


Figure 8.18: Estimated number of true interactions in 2012 data, assuming different values of minimum-bias cross section. The central value is 69.3 mb (solid circles).

but the effect on the result is negligible.

Jet energy scale uncertainty causes an uncertainty on the normalization of 5.5% on background and 1-4.3% for the signal, depending on m_H. Table 8.10 shows the up/down variation on the final yields due to the uncertainty on the jet energy scale, for different mass hypothesis. The effect on background is dependent on m_{llqq}, being negligible at low masses, and reaching a 4% at 600 GeV. The left plot of Fig. 8.19 shows the variations on the Z+Jets distribution due to the jet energy scale uncertainty. The right plot of Fig. 8.19 shows the relative value of these differences as a function of m_{llqq}.

m_HGeV	JES $+1\sigma$ (%)	JES -1σ(%)
230	4.3	-4.2
300	1.3	-1.3
400	1.2	-1.2
600	-0.8	0.9

Table 8.10: Signal efficiency changes due to systematic uncertainties on the jet energy scale.

8.6.5 b-hadron identification

Data-to-MC scale factors (SF_b) are calculated as function of p_T and η to correct the MC for the different efficiency observed in MC with respect to data. Mistag rate scale

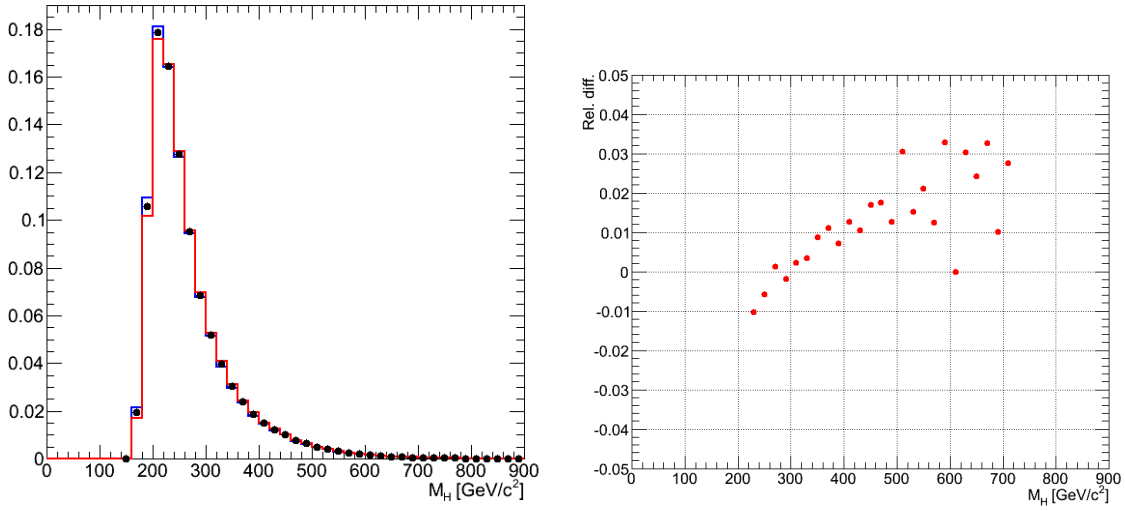


Figure 8.19: Left: shape variation of the m_{llqq} distribution for Z+jets simulated events when varying the systematic uncertainties. Black dots denote the reference shapes. Red and blue histograms indicate the up and down variations of the corresponding systematic uncertainties. Right: relative difference on the shape of the m_{llqq} distribution for Z+jets simulated events when varying the systematic uncertainties.

factors (SF_{mistag}) are calculated to account for light jets reconstructed as b-hadron jets, as well. The associated systematic effect on the analysis is computed by simultaneously varying up and down both scale factors (SF_b and SF_{mistag}) within its associated errors and observing the variation in the final yields.

The effect of SF_b on the background normalization is of 0.4%, 0.8% and 4.5% for the 0-, 1- and 2-btag categories respectively, while the uncertainty of SF_{mistag} introduces a systematic uncertainty of 1.9%, 7.8% and 6.2% for the 0-, 1- and 2-btag categories respectively.

The systematic uncertainty on the signal are listed in Tables 8.11 (muon channel) and 8.12 (electron channel), for various m_H values, for the three btag categories.

8.6.6 Missing transverse energy

The dominant effects are from the knowledge of the rest of the event, such as jet energy reconstruction and pileup. Therefore, the MET uncertainty is covered to a large extent by the previous uncertainties considered. Additionally we investigated how much the MET rescaling procedure described in section 8.3.4 affects the signal selection efficiency, by counting the number of signal events migrating over the MET threshold due to the scaling procedure. The requirement on the MET significance translates thus into about 0.5% uncertainty on the signal final efficiency, and a negligible impact on the background efficiency.

m_H(GeV)	0-btag		1-btag		2-btag	
	$1\sigma_+$ (%)	$1\sigma_-$ (%)	$1\sigma_+$ (%)	$1\sigma_-$ (%)	$1\sigma_+$ (%)	$1\sigma_-$ (%)
200	-3.2	1.5	5.8	-1.3	4.0	-6.7
210	-3.4	1.2	4.8	-1.9	8.5	-2.5
220	-3.1	1.7	4.5	-1.7	6.8	-6.1
230	-3.4	1.8	4.8	-2.1	7.4	-5.3
250	-2.9	2.1	3.8	-2.4	6.3	-5.5
275	-3.4	1.5	4.5	-1.0	6.1	-5.5
300	-3.6	1.4	4.6	-0.69	6.8	-5.7
350	-3.8	1.7	4.7	-0.84	6.7	-6.1
375	-3.7	1.7	5.0	-0.31	5.2	-7.5
400	-3.9	1.6	4.5	-0.74	6.8	-5.7
425	-4.1	1.6	5.1	-0.59	5.9	-5.8
450	-4.1	1.8	4.0	-1.1	8.1	-5.0
475	-3.9	1.8	4.1	-0.6	7.0	-6.5
500	-3.6	2.1	3.6	-0.97	7.1	-6.7
525	-4.1	2.0	4.8	-0.65	6.4	-7.1
550	-3.6	1.9	3.5	-0.87	6.9	-5.6
575	-4.4	1.8	4.9	-0.78	6.7	-5.7
600	-4.3	1.9	4.4	-1.3	7.8	-4.9

Table 8.11: Systematic uncertainty on the signal in the muon channel from heavy quark flavour tagging.

8.6.7 Higgs production mechanism, cross section and branching ratio

The expected kinematics of the Higgs production is subject to uncertainties due to limited knowledge of the underlying PDFs, as well as the shortcomings in the theoretical prediction (missing higher orders in the perturbation series). These uncertainties are propagated to an uncertainty on the selection acceptance and efficiency. The PDF uncertainties are evaluated according to the PDF4LHC recommendations, by evaluating the selection efficiency for the PDF sets CTEQ6 [91], MSTW2008NLO [98] and NNPDF2.1 [99] and their error sets, summarized in Tab. 8.13. The envelope of the various PDF sets is used as the total uncertainty, as recommended, and amounts to 1-4%. The uncertainty noticeably increases for very high Higgs masses.

Uncertainties associated to the appropriate signal modelling described in section 8.5 via the CPS and the interference effects are evaluated. They affect the global efficiency of the selection and, due to the mass-dependence of the efficiency, the shape of the m_{llqq} distribution. The fact that the interference has been computed only at LO while the signal is known at NNLO is also included. The approach proposed in [96, 100] is followed in this case. The uncertainty is negligible below 400 GeV and rises to 3%

m_H(GeV)	0-btag		1-btag		2-btag	
	$1\sigma_+$ (%)	$1\sigma_-$ (%)	$1\sigma_+$ (%)	$1\sigma_-$ (%)	$1\sigma_+$ (%)	$1\sigma_-$ (%)
200	-3.4	1.8	3.9	-3.7	12	-0.18
210	-3.5	1.2	5.4	-1.1	6.9	-5.5
220	-2.9	1.9	4.1	-2.0	7.0	-7.2
230	-3.6	1.4	4.9	-1.9	8.7	-3.5
250	-2.9	2.2	4.0	-2.4	5.1	-6.2
275	-3.5	1.5	4.7	-1.3	6.3	-4.6
300	-3.9	1.5	5.9	-0.26	5.9	-7.7
350	-3.9	1.5	4.7	-0.41	7.1	-6.5
375	-3.9	1.8	4.5	-1.1	7.5	-5.7
400	-3.9	1.9	4.0	-1.1	8.2	-6.0
425	-4.0	1.8	4.8	-0.67	6.9	-6.6
450	-3.9	1.5	3.8	-0.43	7.8	-5.3
475	-4.3	1.8	5.2	-0.54	6.1	-6.2
500	-4.1	2.0	5.2	-0.83	5.8	-6.7
525	-4.0	2.0	3.8	-1.3	8.1	-5.5
550	-4.0	1.8	4.5	-1.1	6.2	-5.1
575	-4.0	1.9	4.1	-0.67	7.1	-6.2
600	-4.4	1.8	5.2	-0.68	6.7	-6.1

Table 8.12: Systematic uncertainty on the signal in the electron channel from heavy quark flavour tagging.

at 600 GeV, with only small dependence on b-tag category. Additionally the line-shape used is re-extracted with the alternative line shape models, which are depicted in Fig. 8.20. The tail caused by mismatched jets is not affected at all as it is a random mixture of events, averaging out any shifts from the uncertainty. The core of the signal distribution is only weakly affected by the uncertainty. In the worst case (the highest mass we consider), the peak-position shifts by 2 GeV (compared to a sigma of 60 GeV) and the sigma changes by 1 GeV.

The Higgs production cross-section uncertainty depends on production mechanism, either gluon fusion or vector boson fusion. The gluon fusion mechanism drives the total uncertainties as it dominates over the VBF, but still gg and VBF errors are estimated separately and for each mass point according to the LHC Higgs Working Group prescription [92]. The total weighted error is in the range 13−15%.

It must be noted that this uncertainty is relevant only for the study of the ratio to SM expectation R, while it does not affect the results related to the absolute cross-section.

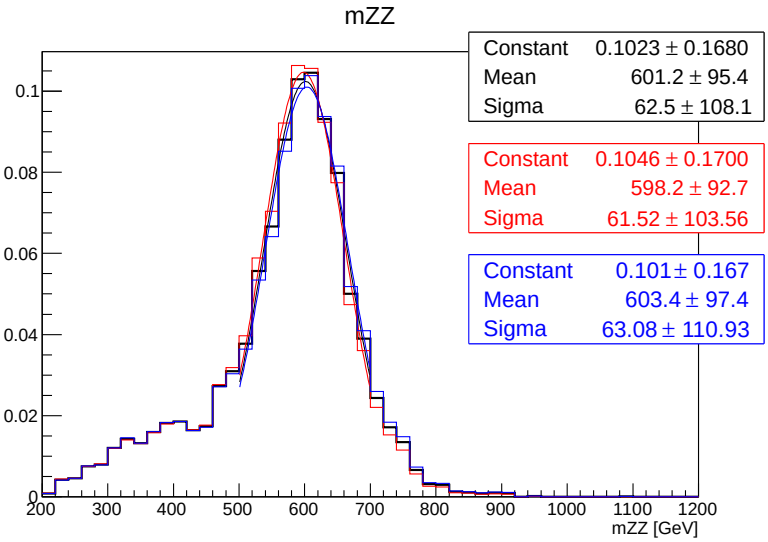


Figure 8.20: Reconstructed m_H= 600 GeV Higgs signal (area normalized) with the nominal lineshape (black) and systematic variations (blue/red). Gaussian fits to the core of the distribution are overlaid.

8.6.8 p_T-dependent correction to the Z+jets simulation

The uncertainty associated to the correction applied to the Z+jets simulation based on the differences of the p_T distribution of the $llqq$ object, described in section 8.4, is evaluated comparing the m_{llqq} distribution with and without the correction. A mass-dependent systematic uncertainty is obtained as the difference of those distributions, and goes up to 3% at high m_{llqq} values.

8.6.9 Data to MC residual differences on the Z+Jets lineshape

Residual differences between data and MC in the sidebands are used to generate templates to quote for additional systematic uncertainty on the Z+Jets predicted m_{llqq} shape in the signal region. They are shown in the upper plots of Fig. 8.21

	m_H(GeV)				
PDF	200	400	600	800	1000
CTEQ66	$^{+0.6}_{-0.7}$	$^{+0.8}_{-1.0}$	$^{+0.8}_{-1.1}$	$^{+1.5}_{-2.0}$	$^{+2.6}_{-3.2}$
MSTW2008NLO	$^{-0.2}_{-0.5}$	$^{+0.6}_{+0.2}$	$^{+0.8}_{+0.4}$	$^{+1.5}_{+0.7}$	$^{+2.5}_{+1.2}$
NNPDF2.1	$^{+0.8}_{+0.2}$	$^{+1.4}_{+0.75}$	$^{+1.5}_{+0.9}$	$^{+2.7}_{+1.4}$	$^{+4.3}_{-2.4}$
Total	$^{+0.8}_{-0.7}$	$^{+1.4}_{-0.8}$	$^{+1.5}_{-1.1}$	$^{+2.7}_{-2.0}$	$^{+4.3}_{-3.2}$

Table 8.13: Systematic uncertainties on the signal acceptance following PDF4LHC recommendations.

independently for the six channels.

A smooth parametrization to cover fluctuations in the residuals is performed, which is a more conservative approach than a straight linear fit to the points.

The parametrization is used to produce alternative Z+jets lineshapes that will be included in the Higgs presence test hypothesis described in next chapter. Plots in Fig. 8.21 show the alternative lineshapes for the six exclusive channels.

8.7 Results

The m_{llqq} distribution for the six exclusive channels (the two lepton categories times the three b-tag categories) is depicted in Fig. 8.22 and Fig. 8.23 for the m_{jj} sideband region. After all the analysis strategy has been decided, the m_{llqq} distribution in the signal region is presented in Fig. 8.24 and Fig. 8.25 in linear and logarithmic scale. The histograms are plotted with a m_{llqq}-dependent binning, to account for the low statistics in the high-mass regime, and avoid large statistical fluctuations. The systematic uncertainty in the background estimation, detailed in the previous section, are not shown in the figures.

The normalization of the simulated background (Z+jets and dibosons) is free to float in the signal region, with the constraint of the number of events in the m_{jj} sidebands.

The background prediction reproduces very well the observed data, but a more accurate interpretation of the results is presented in chapter 10.

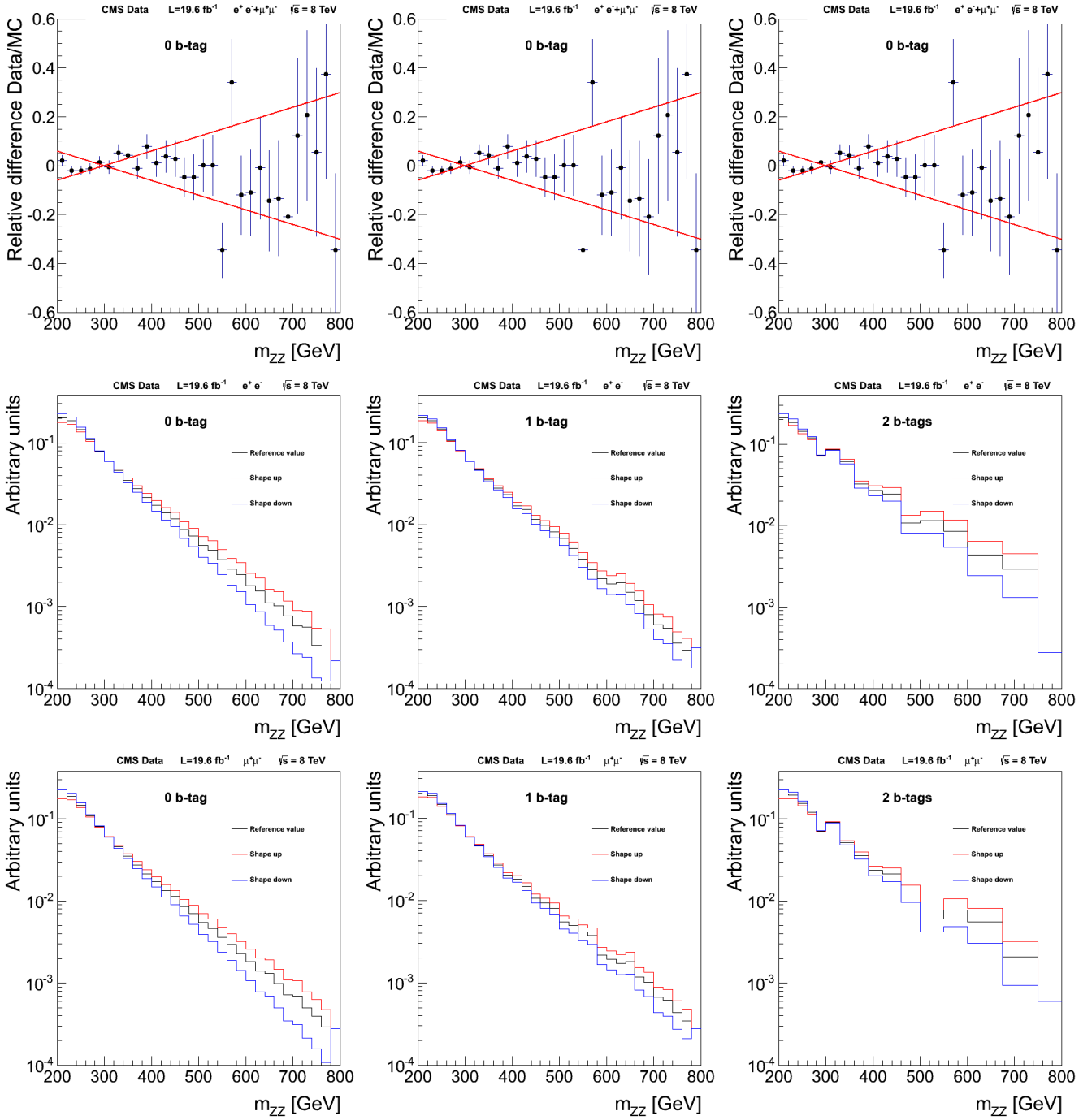


Figure 8.21: Residual differences in the m_{llqq} distributions between the data and the background, in the M_{jj} sideband control region (top). Alternative templates for the background prediction taking into account those residual variations, for the electron (middle) and the muon channels (bottom).

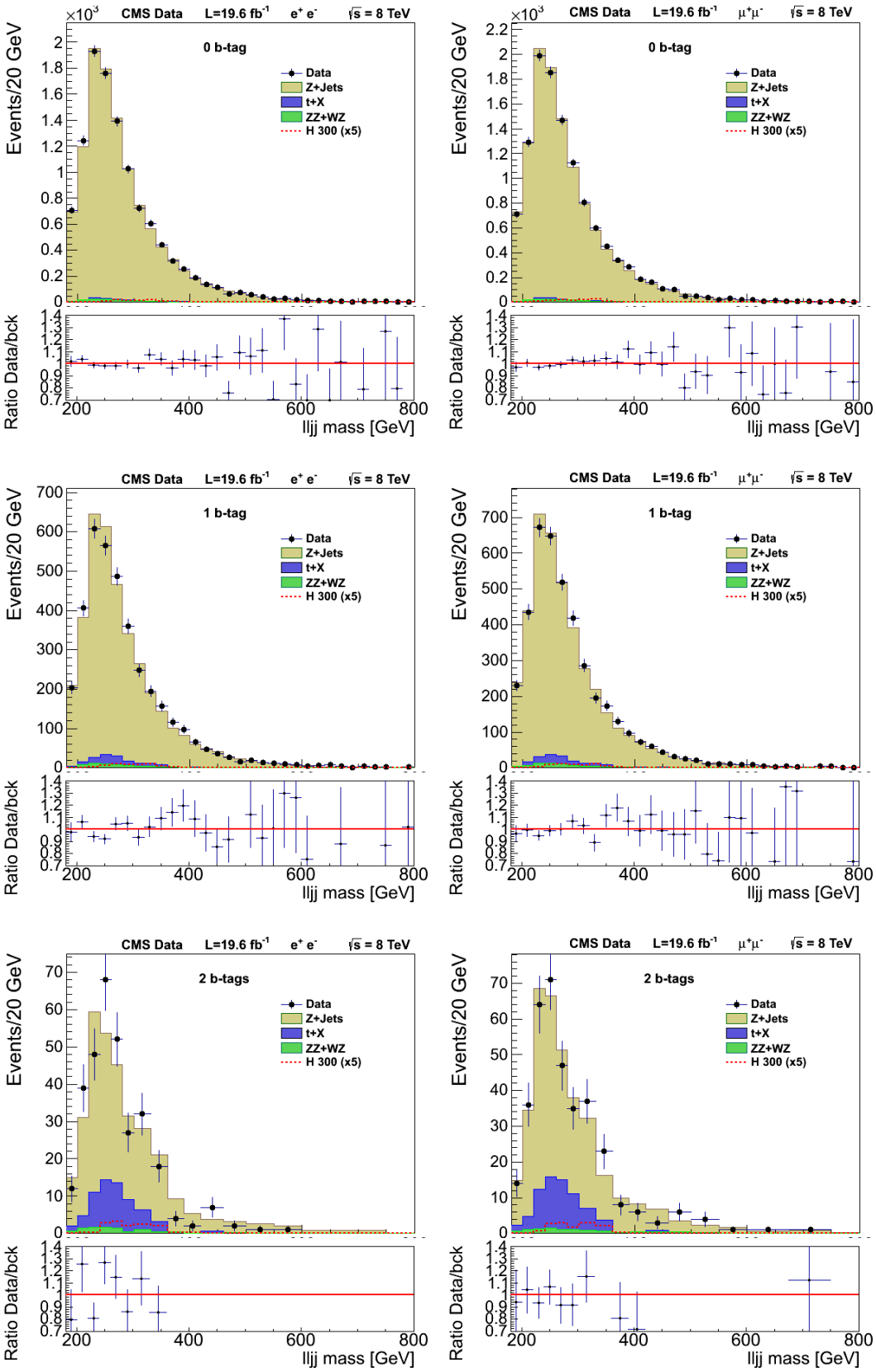


Figure 8.22: Mass distributions of the $lljj$ system for events in the m_{jj} sideband region in the electron (left) and muon (right) channels. From top to bottom, plots correspond to the 0-, 1-, and 2-btag categories. In the 2-btag category a variable bin size is used to compensate for the low number of events in the tail of the distributions. The dots are data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue $t\bar{t}$ events from data.

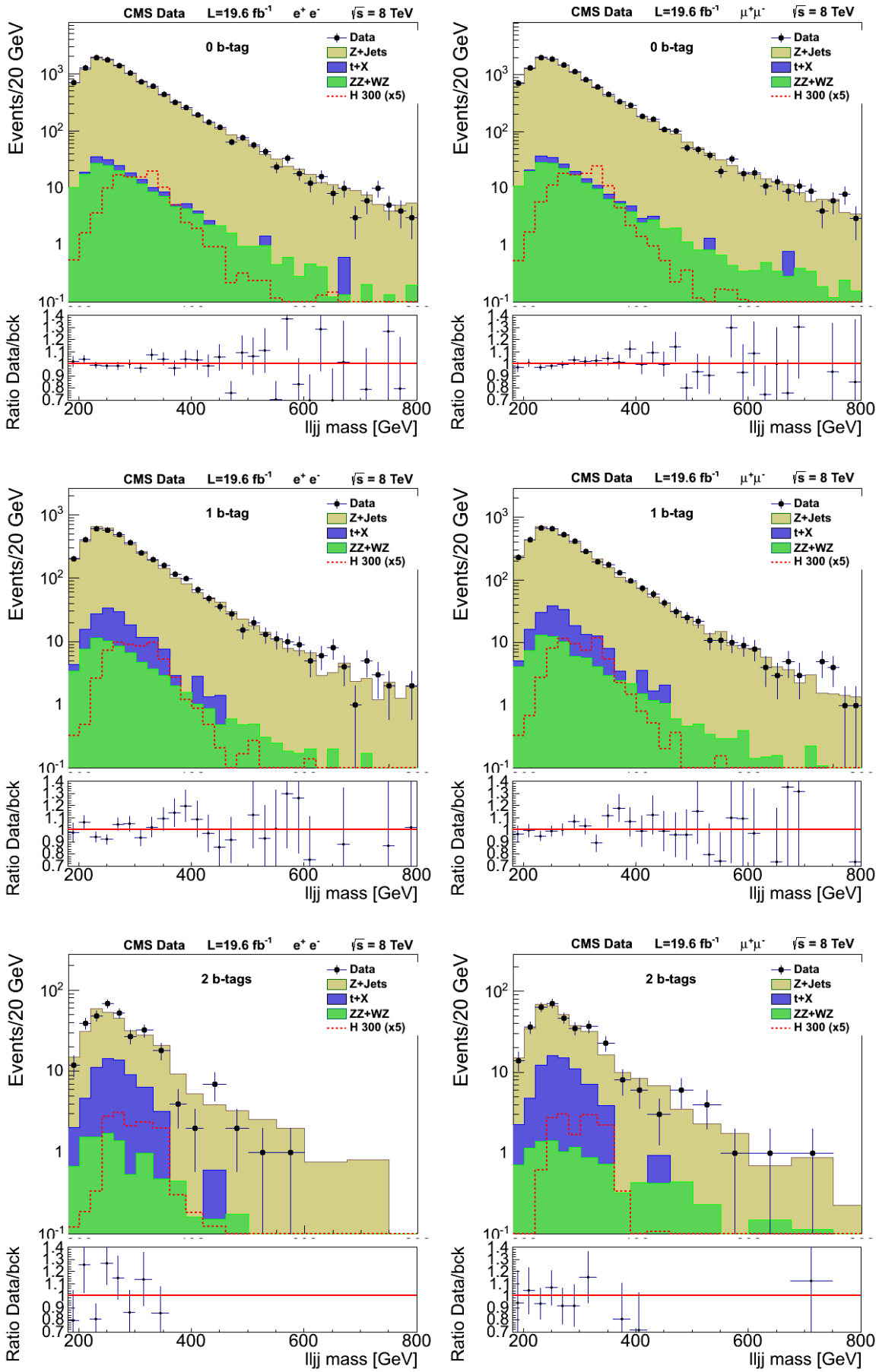


Figure 8.23: Mass distributions of Fig. 8.22 in logarithmic scale.

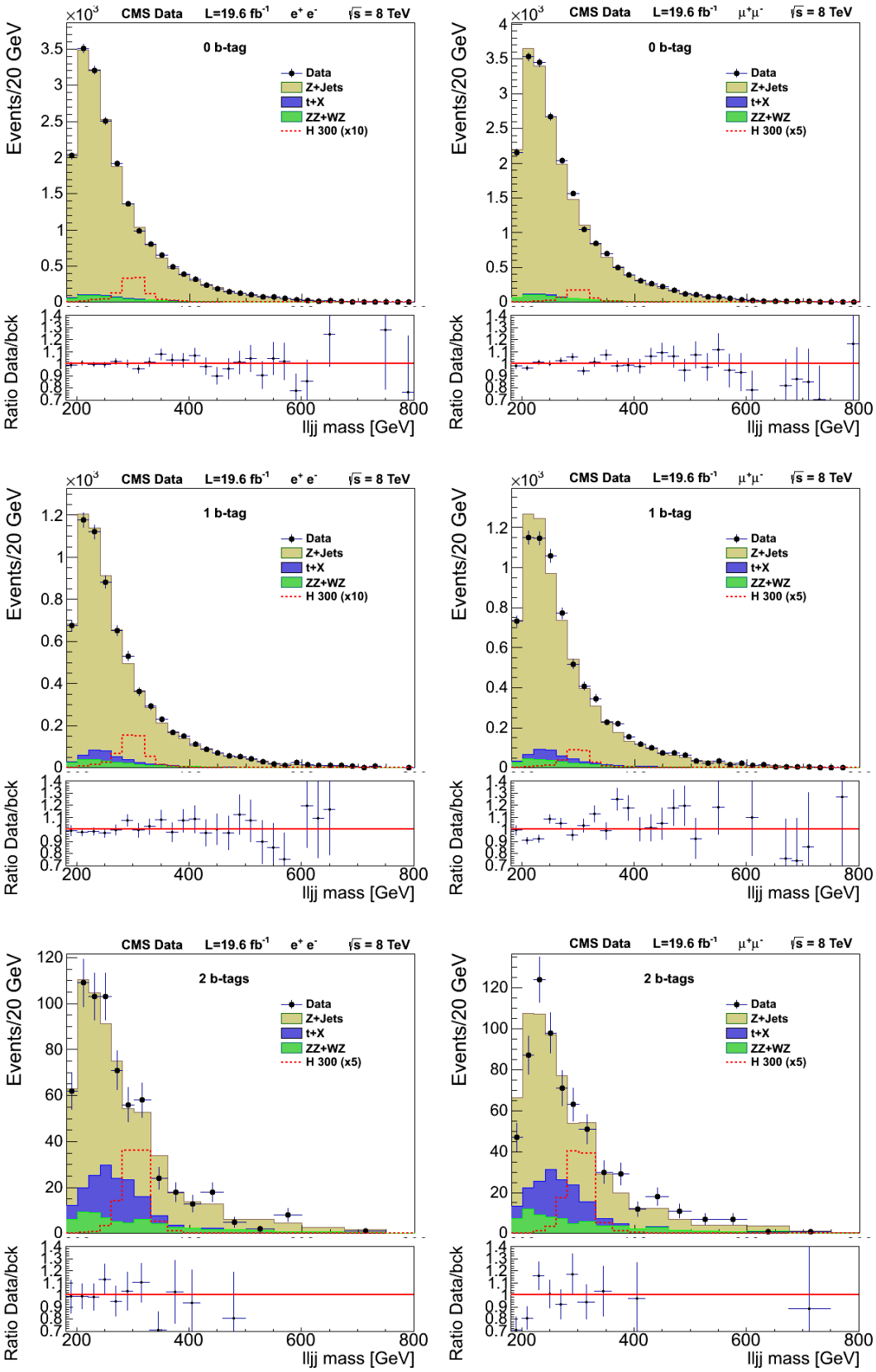


Figure 8.24: Mass distributions of the $lljj$ system for events in the signal region in the electron (left) and muon (right) channels. From top to bottom, plots correspond to the 0-, 1-, and 2-btag categories. The dots are data, the ochre histogram indicates the corrected Z+jets simulation, the green histogram simulated diboson background and the dark blue $t\bar{t}$ events from data. The systematic uncertainty on estimation of the background, as described in section 8.6, is not shown in the histograms.

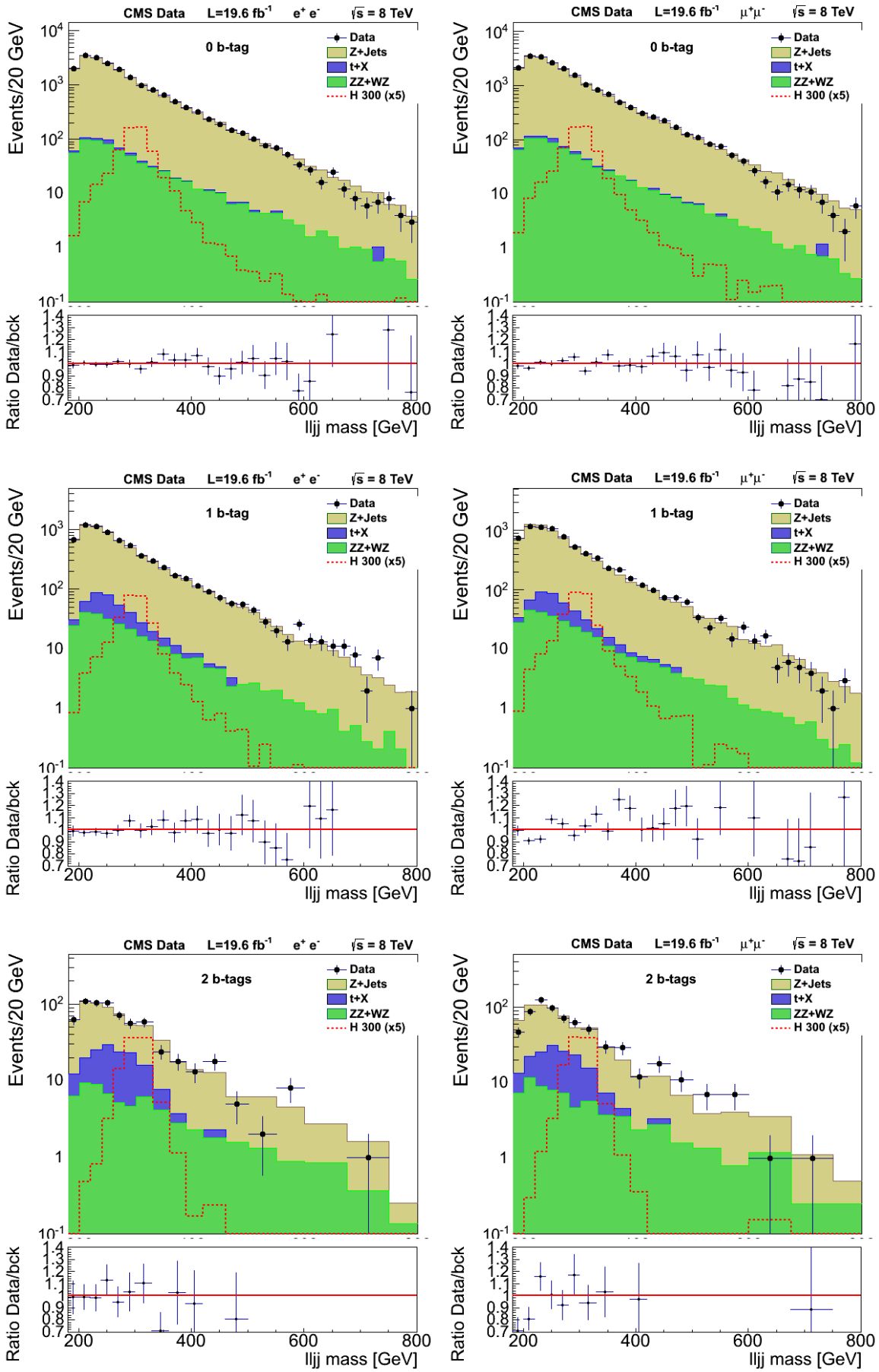


Figure 8.25: Mass distributions of Fig. 8.24 in logarithmic scale.

Chapter 9

Search for the Higgs boson in the $H \to ZZ \to l^+l^-q\bar{q}$ decay channel in pp collisions at $\sqrt{s}$= 7 TeV

The first analysis by CMS on the $H \to ZZ \to l^+l^-q\bar{q}$ final state was published in 2012 [4] and updated few months later [6], with the data recorded during 2011, which corresponds to an integrated luminosity of $\mathcal{L}$=5.1 fb^{-1}, from collisions at $\sqrt{s}$=7 TeV in the center of mass. The features of the $H \to ZZ \to l^+l^-q\bar{q}$ analysis have been already described in chapter 8. The analysis developed during 2011 followed a similar approach as one for the 8 TeV dataset described in chapter 8. However, some differences are present, due to the improvements achieved in the generated simulations, the detector performance and the development of analysis techniques, as well as the variations in the data conditions (pile-up, size of the samples ...). The present section reports the main differences between both analysis.

9.1 Event selection

There are several differences in the event selection, although they have, in most of the cases, a negligible impact on the final selection:

- A combined (detector-based) isolation is used in leptons, instead of PF isolation. The pile-up contribution substracted to the measured energy in the isolation cone was performed using "effective areas" technique in both muons and electrons, instead of only electrons. Nevertheless, the effect of this change was negligible.
- The pile-up contamination during 2011 was lower than in 2012, and hence the quality cuts, including the β threshold, applied to both jets during 2012 were not demanded.
- An additional discriminant was used in order to reduce, on average, contribution of jets originated from gluon splitting instead of quark hadronization, as only

the latter are result of a Higgs boson decay. This discriminant was applied to the 0 b-tag category, where light jet production dominates, and tried to use topological differences between quark and gluon jets. In particular: the number of charged hadronic particle tracks, the number of photons and neutral hadrons and the quantity $\sqrt{\frac{\sum p_T^2}{(\sum p_T)^2}}$ added over all jet constituents, as gluon jets tend to be softer and with more constituents. The systematic uncertainty associated to this discriminant could not be updated during the 2012 analysis, so it was decided not to make use of it, taking into account that the increase of luminosity reduced the statistical uncertainty, so the 2 b-tag category became the most sensitive for Higgs searches, dominating over the 0 b-tag sensitivity, contrary to the 2011 case, when the 0 b-tag was the most sensitive.

- The angular LD discriminant was optimized as a function of the m_{llqq}, independently for the 3 b-tag categories. During 2012 analysis this optimization was repeated, and a negligible improvement was obtained for a variable threshold compared to a constant one, so the latter was chosen in order to simplify the analysis.

- Since the main limitation in the resolution of the m_{llqq} distribution is the jet resolution, the fact that the two jets are coming from a Z boson was used to perform a kinematic fit to constrain the m_{jj} mass to the nominal Z mass and modify the m_{llqq} distribution accordingly,

- the dilepton and dijet windows were slightly different, optimized according to the present resolution, efficiency and background determination strategy. The dilepton cut was $70<m_{ll}<110$ GeV and the dijet cut was $75<m_{jj}<105$ GeV.

- The $\not{E}_T$ significance cut was only applied to the 2 b-tag category. Its intention is to remove $t\bar{t}$ background, present in categories with b-jets, but in order to unify criteria in all categories it was extended to the other two categories during 2012. The effect is negligible.

- The Track Counting High Efficiency b-tag algorithm, mentioned in section 7.3.1, was used, accordingly to the recommendations of the B-tag Working Group of CMS, and as the one giving the best possible performance.

9.2 Background determination

The background determination has suffered several changes. During 2011, the data-driven top background determination was not ready, so the top simulation was used instead.

The background was determined globally for all background processes together: Z+Jets, $t\bar{t}$ and dibosons, instead of separately for each process as in 2012. It was determined from the MC background simulation and from the data in the sidebands,

as for the Z+Jets in 2012. In particular, the background m_{llqq} distribution in the signal region $N_{\mathrm{SR}}(m_{ZZ})$ was obtained from

$$N_{\mathrm{SR}}(m_{ZZ}) = N_{\mathrm{SB}}^{data}(m_{ZZ}) \times \frac{N_{\mathrm{SR}}^{\mathrm{MC}}(m_{ZZ})}{N_{\mathrm{SB}}^{\mathrm{MC}}(m_{ZZ})} \tag{9.1}$$

In conclusion, a reweight is performed on the m_{llqq} ratio instead of in the $\mathrm{p}_{\mathrm{T}}^{llqq}$ ratio. The normalization of the background is also directly determined from the data in the sidebands and the SR/SB ratio in the MC, whereas in the 2012 analysis the normalization is free to be float in the fit of the limit calculation, with the constraint from the sideband relative normalization.

9.3 Statistical Interpretation of the results

During the 2011 analysis, a very important effort was put to extend the m_H range under study as wide as possible, even below the on-shell mass of the ZZ pair. The range of study was, in turn, 183-600 GeV, even when the rapid turn-on of the background in the region of low mass is very difficult to model. An additional analysis searching the range 120-170 GeV was developed, but with very low sensitivity.

An unbinned shape analysis was performed in order to obtain exclusion limits. Both background and shape distribution were parametrized.

The background distribution was fitted to the convolution of a "Crystal Ball" times Fermi function. A Crystal ball function is a Gaussian bell with a power-law tail. The Fermi function follows the Fermi distribution $\frac{1}{e^{(\alpha-x)/\beta}+1}$.

In Fig. 9.1 the final m_{llqq} invariant mass distributions are shown for the 3 b-tag categories, with no distinction of the lepton flavour. The background expectation corresponds to the blue line, whereas the histograms only illustrates the prediction obtained by the simulation.

The signal distribution was splitted in two subsamples in order to be correctly parametrized: matched candidates and unmatched. The former are composed of those where the two selected jets match the two jets from the Higgs decay at generator level. It is fitted to a double crystal ball (similar to a single crystal ball but with two power-law tails). The unmatched subsample corresponds to those candidates where the selected jets are not the correct ones at generator level, and will produce a smooth broader distribution, fitted to a single crystal ball convoluted with an additional triangle function.

The exclusion limits obtained in this analysis are included in the chapter 10.

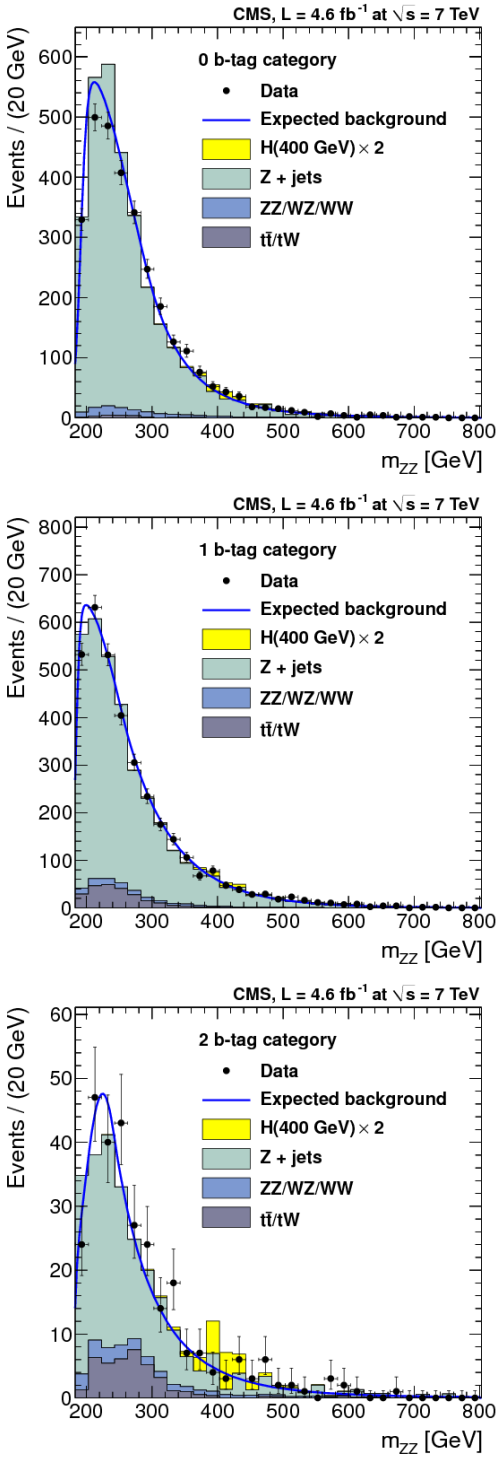


Figure 9.1: From top to bottom, the m_{llqq} invariant mass distribution after final selection in the 0,1 and 2 b-tag category. Points with error bars show distributions of data and solid curved lines show the prediction of background from the sideband extrapolation procedure. For illustration solid histograms depict the background expectation from simulated events with the different components illustrated. Also shown is a hypothetical signal with the mass of 400 GeV and cross section 2 times that of the Higgs boson, which roughly corresponds to expected sensitivity.

Chapter 10

Statistical Interpretation

The statistical inference allows statements about evidence or exclusion of a signal from the outcome of the performed analysis. The present chapter depicts the statistical analysis performed on the events selected in chapter 8, to establish whether any deviation is compatible with a fluctuation of the background or more likely, may be interpreted as signal evidence, as well as establish upper limits to the signal production in the LHC.

The m_{llqq} invariant mass distribution is used as a discriminant of the background only hypothesis versus the signal+background hypothesis. The range of Higgs masses hypothesis under study covers $230<m_H<600$ GeV, although the input m_{llqq} distributions cover the range 220-800 GeV (Figs. 8.24 and 8.25). As no excesses have been observed in those figures nor in the equivalent of the 7 TeV analysis (Fig. 9.1), upper limits to the production of a SM Higgs boson in the LHC will be presented for both 7 TeV and 8 TeV dataset, and its combination.

10.1 Statistical analysis

The statistical methodology for the analysis of the Higgs boson searches at the LHC has been developed by the ATLAS and CMS collaborations through the LHC Higgs Combination Group [101].

It is constructed over several quantities:

- The observable measured in the data. In our case, it is given by the m_{ZZ} distribution.
- The models for background only and signal + background hypothesis, expressed as probability density functions

$$P(obs|b) \qquad\qquad P(obs|s+b) \tag{10.1}$$

The p.d.f. for the 7 TeV dataset analysis are unbinned, given by the parametrization expressed in section 9.3, whereas in the 8 TeV analysis the input models are the binned histograms shown in Fig. 8.24

- The systematic uncertainties referred in section 8.6. A nuisance parameter θ_i is assigned to each of the systematic sources, accounting for correct correlations among them. Both signal and background are therefore expressed as functions of them: $s(\theta_i)$ and $b(\theta_i)$. Most of the nuisance parameters θ_i express deviations from the input central value in independent measurements. This information is contained in the p.d.f $p(\tilde{\theta}_i|\theta_i)$ which expresses the probability of measuring $\tilde{\theta}_i$ given its true value θ_i. Some nuisance parameters may be not constrained by $p(\tilde{\theta}_i|\theta_i)$. For example, in the case of the Z+jets normalization in the 8 TeV analysis, there is no the nuisance parameter associated to it. Consequently, it is free to float in the fit, only constrained by a nuisance parameter containing the information of the yields in the sideband region.

When combining results of different periods (7TeV + 8TeV) common uncertainties are considered fully and positive correlated, except for statistical uncertainties, or when methods to calculate them were very different. In that case, no correlation is assumed.

A test of hypothesis is performed independently for each m_H mass hypothesis considered. A grid of m_H points is defined to map the range under consideration.

The signal strength $\mu = \sigma/\sigma_{SM}$ is defined as the ratio of the predicted signal cross section with respect to the SM case, which allows a more general interpretation, and an easy accommodation in case unknown phenomenology or New Physics processes alter the production rates. In consequence, the signal model is often expressed as $\mu \cdot s(\theta_i)$.

The likelihood function is built from the product of likelihoods of the data in each of the six channels times the product of the probability densities for the measurements associated to the nuisance parameters $\tilde{\theta}_i$

$$L(data, \tilde{\theta}|\mu, \theta) = \prod_{c=1}^{N_{ch}} L_c(data_c|\mu, \theta) \times \prod_{i=1}^{N_\theta} p_i(\tilde{\theta}_i|\theta_i) \tag{10.2}$$

In the case of binned data, like in the 8 TeV dataset analysis, the first term of Eq. 10.2 is a Poisson distribution for each bin, built from the number of observed events N_k^{obs}, and the expected signal and background yields $s_k(\theta)$ and $b_k(\theta)$.

$$L(obs|\mu, \theta) = \prod_{k=1}^{N_{Bins}} Poisson(N_k^{obs}, \mu \cdot s_k(\theta) + b_k(\theta)) \tag{10.3}$$

In the case of unbinned data, like in the 7 TeV dataset analysis, the likelihood is built from a Poissonian from the total observed and expected yields N^{obs}, $s(\theta)$ and $b(\theta)$ and from the normalized signal and background models.

$$L(obs|\mu, \theta) = Poisson(N^{obs}, r \cdot s(\theta) + b(\theta)) \times \prod_{k=1}^{N_{evt}} f(x_k|\mu, \theta) \tag{10.4}$$

where x corresponds to the invariant mass m_{ZZ}, N_{evt} is the number of events and $f(x_k|\mu,\theta)$ is

$$f(x_k|\mu,\theta) = \frac{r \cdot s(\theta)}{\mu \cdot s(\theta) + b(\theta)} f_s(x,\theta) + \frac{b(\theta)}{\mu \cdot s(\theta) + b(\theta)} f_b(x,\theta) \tag{10.5}$$

where $f_b(x,\theta)$, $f_s(x,\theta)$ are the parametrizations of m_{ZZ}.

Both approaches, binned and unbinned, are equivalent in the large bin number limit.

The measurements $\tilde{\theta}_i$ that constrain the nuisance parameters are introduced according to the origin of the uncertainty. Uncertainties in the shape are introduced via alternative shapes with $\pm 1\sigma$ variation with respect the central value. The uncertainties affecting the normalization are mainly introduced as log-normal distributions of the corresponding nuisance parameters. However, other possibilities are allowed. For instance, nuisances representing the statistical uncertainties on the number of events in a control region or simulated sample, are taken to be the observed event count and the expected yield, so that $p_i(\tilde{\theta}_i|\theta_i)$ is a Poisson probability.

The RooStats tool [102] is used to extract the limits.

10.1.1 Exclusion limits

A modified frequentist method CL_s [103, 104] is chosen to establish the significance of excesses and exclusion limits. The test statistics q is a single number that accounts for all the information given from the measurements and expectations enumerated and ranks observations according to whether they are more consistent with $s+b$ or b hypothesis.

The exclusion limits are established from the test statistics q_μ which is the most powerful discriminator according to the Neyman-Pearson lemma:

$$q_\mu = 2log \frac{L\left(data|\mu \cdot s(\hat{\theta_\mu}) + b(\hat{\theta_\mu})\right)}{L\left(data|\hat{\mu} \cdot s(\tilde{\theta}) + b(\tilde{\theta})\right)} \qquad 0 < \hat{\theta_\mu} < \mu \tag{10.6}$$

where the subscript in $\hat{\theta_\mu}$ indicates that the likelihood in the numerator is maximized under the hypothesis of a given signal of strength μ, while $\hat{\mu}$ and $\hat{\theta}$ are the ones which maximize the likelihood at the denominator. In order to force one-sided limits on the signal production rate, we constrain $0 < \hat{\mu} < \mu$.

In a frequentist approach, expected distributions of the test statistics q would be obtained by generating a set of pseudo-dataset from the probability density functions $P(data|\mu,\theta)$ and $p(\tilde{\theta}|\theta)$. The values of the nuisance parameters θ, used for generating the pseudo-datasets, are obtained from a fit to data, maximizing the likelihood L under the b and the $s+b$ hypotheses. However, when the statistics are large enough, like this case, the probability density functions, p.d.f, of the test statistic q_μ, for signal + background and background-only hypotheses, are well defined by an analytical

formula [105]. Therefore, it is possible to get the expected limit with error bands without generating pseudo-experiments.

The confidence in the signal + background hypothesis, CL_{s+b}, is defined by the probability to observe a value of the test statistics equal to or larger than the value observed in the experiment:

$$CL_{s+b} = P(q_\mu \geq q_\mu^{obs} | \mu \cdot s + b) \tag{10.7}$$

Equivalently, the confidence in the background only hypothesis, CL_b is defined as the probability to observe a value larger than the observed under the background only assumption

$$CL_b = P(q_\mu \geq q_\mu^{obs} | b) \tag{10.8}$$

The CL_s ratio is defined as the ratio

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{10.9}$$

Therefore, it is possible to exclude the signal existence if the analysis is sensitive enough. If $CL_s \leq a$ for $\mu = 1$, the signal hypothesis is excluded at the 1-a confidence level. The same way, upper limits on μ are quoted at 1-a confidence level adjusting μ until $CL_s = a$ is reached. Typically in the results presented, $a = 0.05$.

In consequence, the exclusion a SM Higgs boson signal at 95% confidence level (CL) is to be interpreted in this context as the statement that the probability of obtaining an outcome as background-like as the observed one if the signal hypothesis were to be true is 5% or less; in the remaining 95% or more of the experimental outcomes, the presence of the signal would have been more evident than what observed in the data.

10.2 Results

Modified frequentist upper limit plots at 95% CL are presented for the 7 TeV and 8 TeV analysis. Figure 10.1 shows the 7 TeV result. Observed limits are indicated by a solid-dotted line and for comparison the expected limits (dashed line) are shown as well, together with the 68% (green band) and 95% (yellow band) probability ranges around the expected limits.

The production of a SM Higgs boson with masses between $340 < m_H < 400$ GeV is excluded at 95% CL by this analysis alone, which is the region under $\mu = 1$. Above $m_H \sim 480$ GeV an excess of events are observed, but analysis at 8 TeV suggest that it corresponds to a fluctuation of the background. Figure 10.2 depicts the exclusion limits for the 8 TeV analysis with 19.6 fb^{-1} of recorded luminosity. This analysis extends the exclusion range to 275-600 GeV, due to the increase of the sample and the improvements in the analysis.

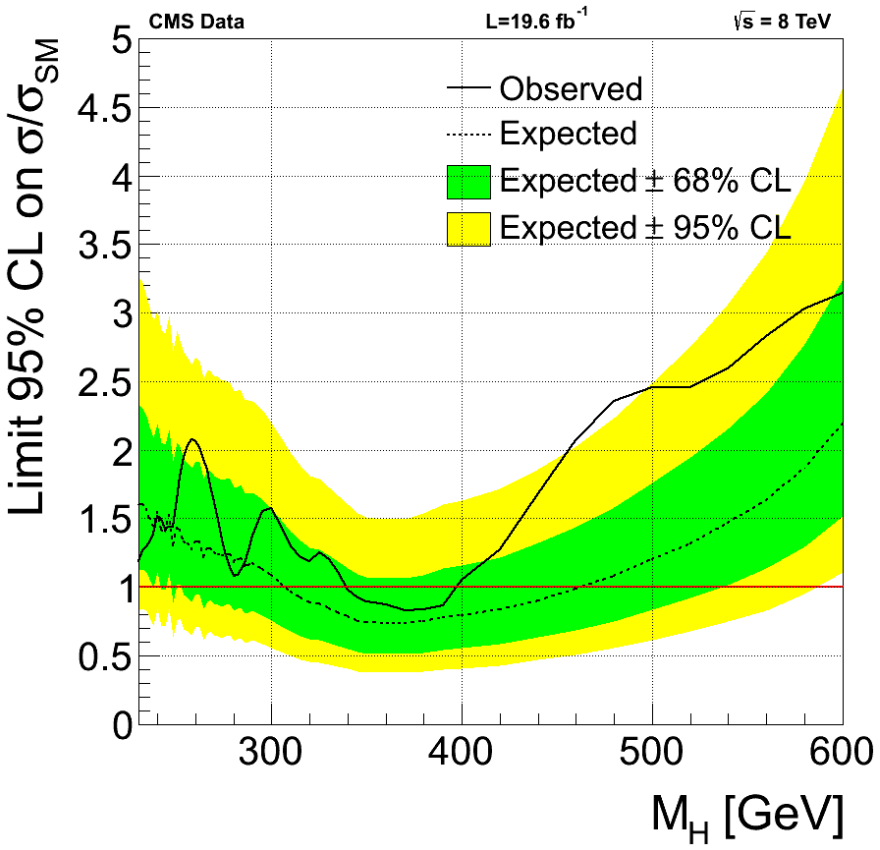


Figure 10.1: Observed (solid) and expected (dashed) 95% CL upper limits on the ratio of the production cross section to the SM expectation for the Higgs boson, for the analysis performed with the 7 TeV dataset, obtained with the modified frequentist CLs technique on its asymptotic limit. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid horizontal line at unity indicates the expectation for a SM-Higgs-like boson.

Figure 10.2 is the combination of the six exclusive channels considered. Exclusion limits for each of the individual channels are shown in Fig. 10.3

Apart from exclusion limits, upper limits on the absolute production cross section times branching fraction of the $l^+l^-q\bar{q}$ final state are shown in Fig. 10.4.

Both analysis (for 7 and 8 TeV dataset) may be combined in the same way. The result is shown in Fig. 10.5. The larger sensitivity of the 8 TeV analysis dominates in the combination.

Finally, the 2l2q final state is one of the several final states exploring searches in the high mass regime. Other ZZ decays have also explored the same models, like the $4l$ or the $2l2\nu$. Figure 10.6 shows the combined exclusion limit of all ZZ searches in the range up to 1 TeV, although the present analysis only contributes up to 600 GeV.

In conclusion, no indication of any signal process have been observed during this analysis. Exclusion limit to the SM Higgs boson hypothesis have been established over

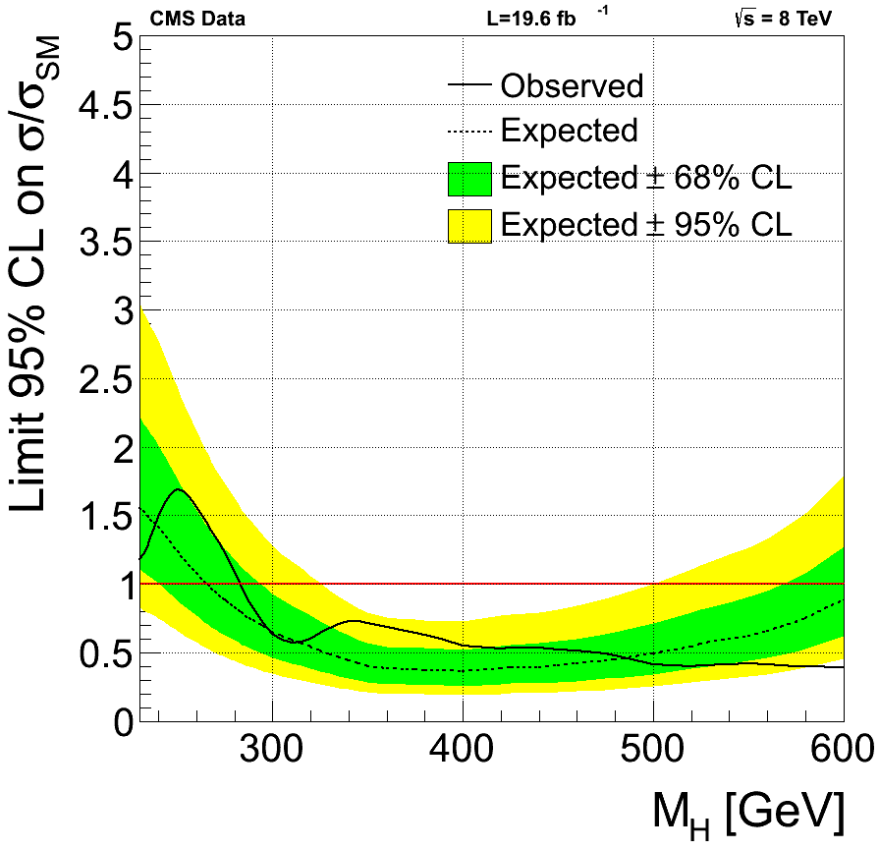


Figure 10.2: Observed (solid) and expected (dashed) 95% CL upper limits on the ratio of the production cross section to the SM expectation for the Higgs boson, for the analysis performed with the 8 TeV dataset, obtained with the modified frequentist CLs technique on its asymptotic limit. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid horizontal line at unity indicates the expectation for a SM-Higgs-like boson.

almost the whole range of study. Exclusion limits to other models of interest, involving new Physics, have not been explicitly established, as it depends of the specific model considered. However, given the present results, no hint or indication of any of them have been observed, implying that the existence of any new Physics in this range might imply a very low production rate.

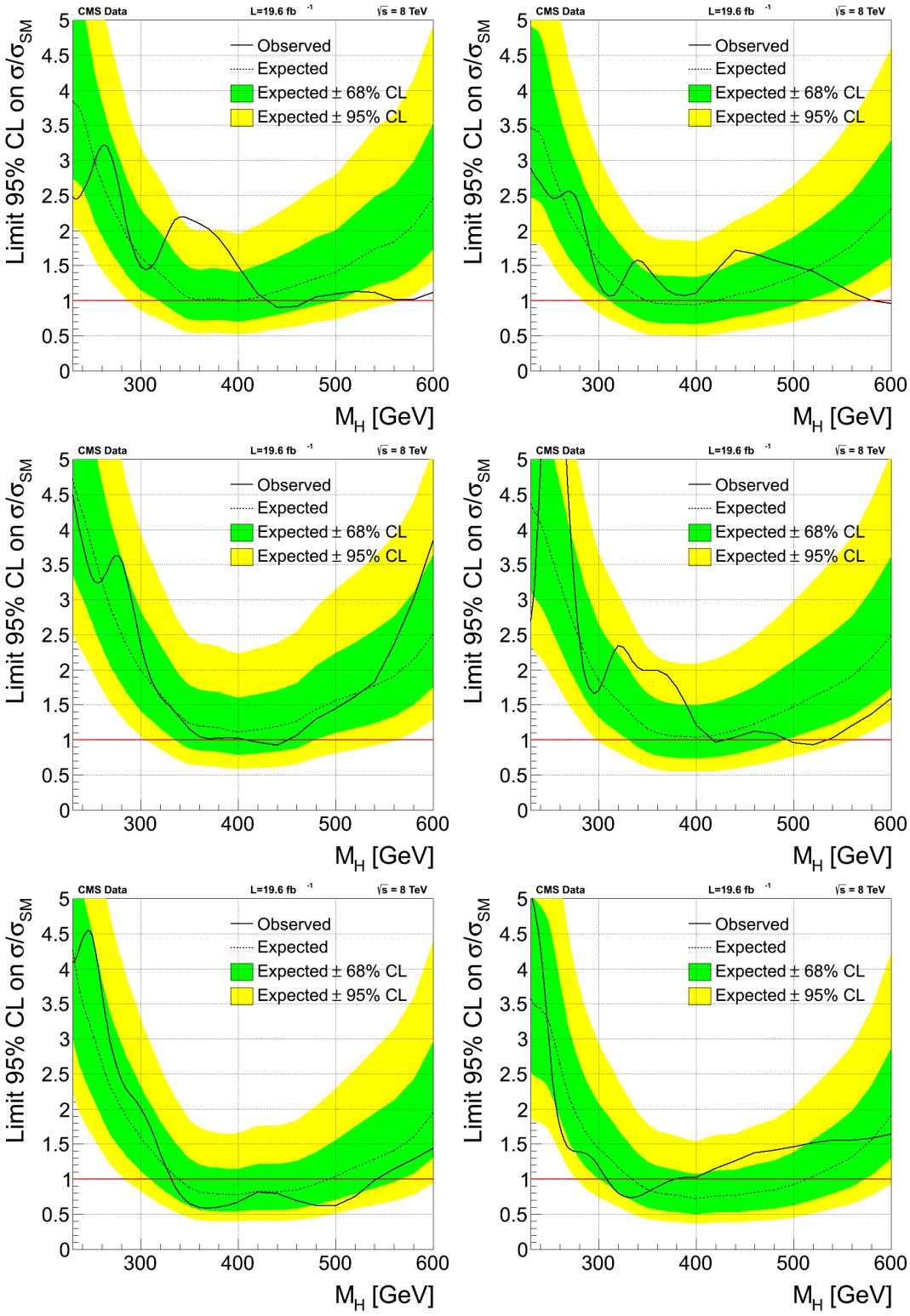


Figure 10.3: From top to bottom, observed (solid) and expected (dashed) 95% CL upper limits for the 0, 1 and 2 b-tag category in the 8 TeV analysis. Left column correspond to the electron channel and right column corresponds to the muon channel.

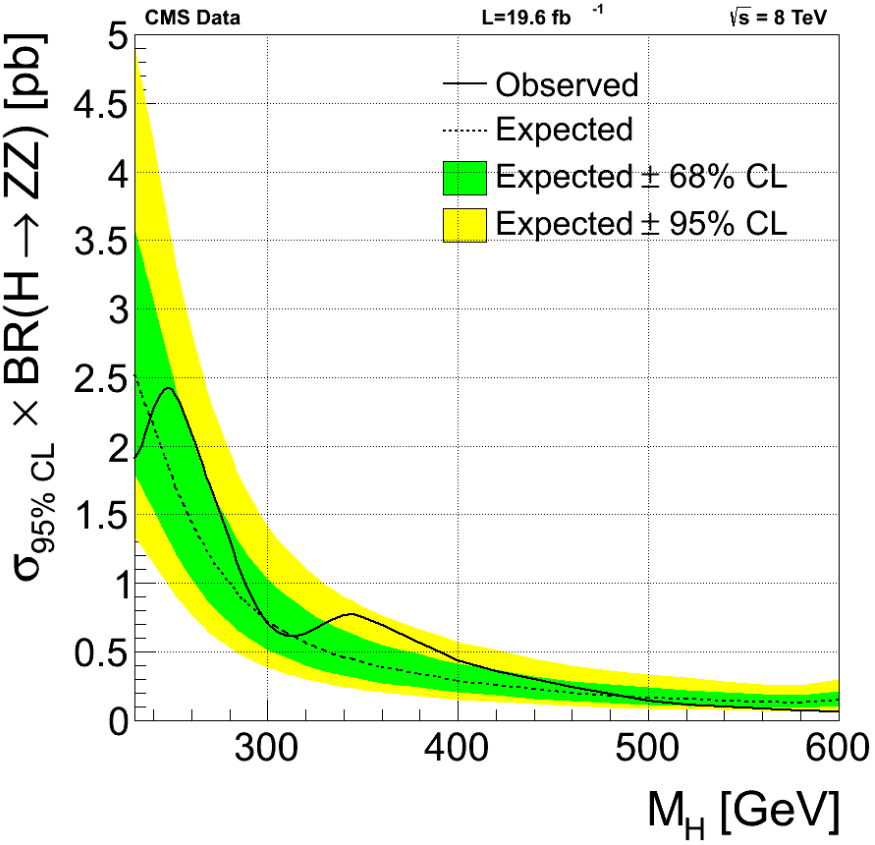


Figure 10.4: Observed (solid) and expected (dashed) 95% CL upper limits on the production cross section times branching fraction of H $\rightarrow$ ZZ, obtained with the modified frequentist CLs technique on its asymptotic limit . The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively.

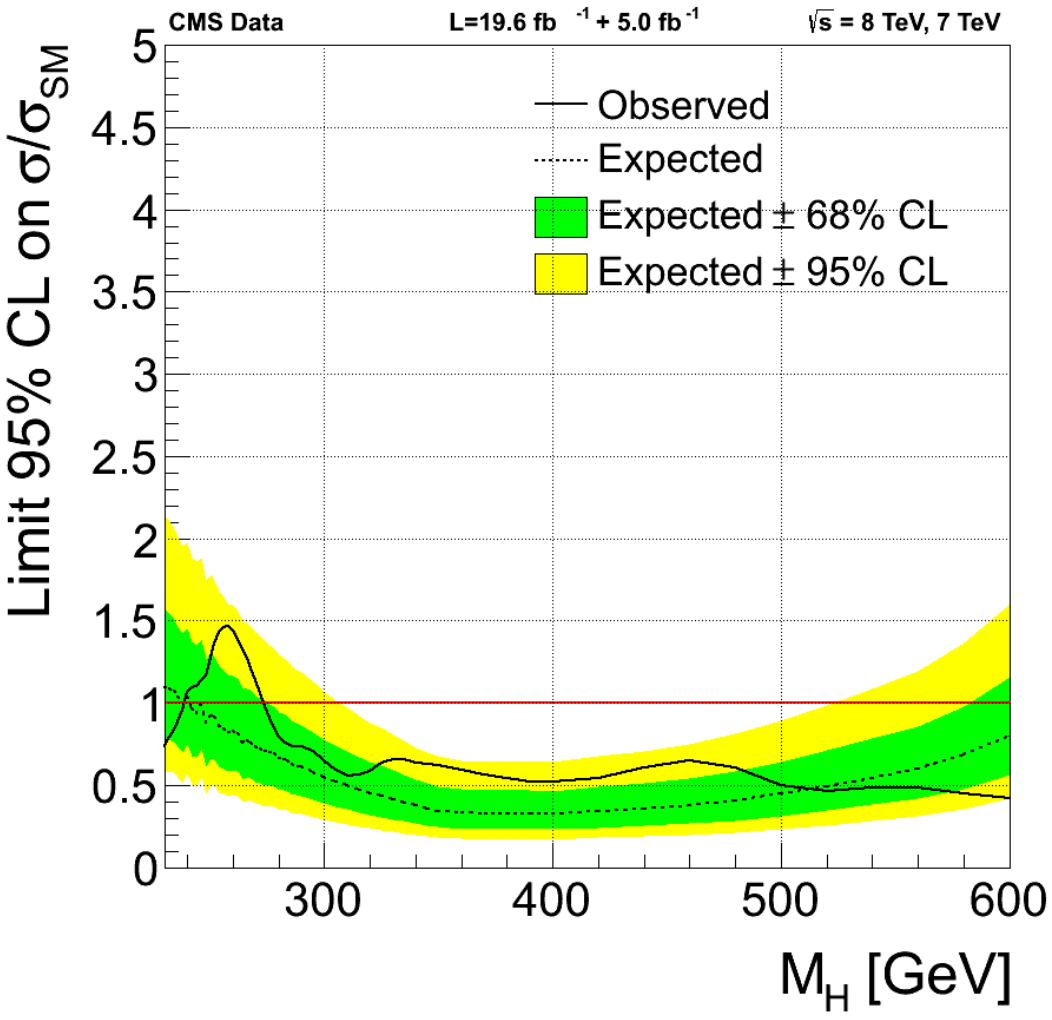


Figure 10.5: Combined limit of theH $\to$ ZZ $\to l^+l^-q\bar{q}$ analysis at 7 TeV (5 fb^{-1}) and 8TeV (19.6 fb^{-1}). Observed (solid) and expected (dashed) 95% CL upper limits on the ratio of the production cross section to the SM expectation for the Higgs boson, . The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid horizontal line at unity indicates the expectation for a SM-Higgs-like boson.

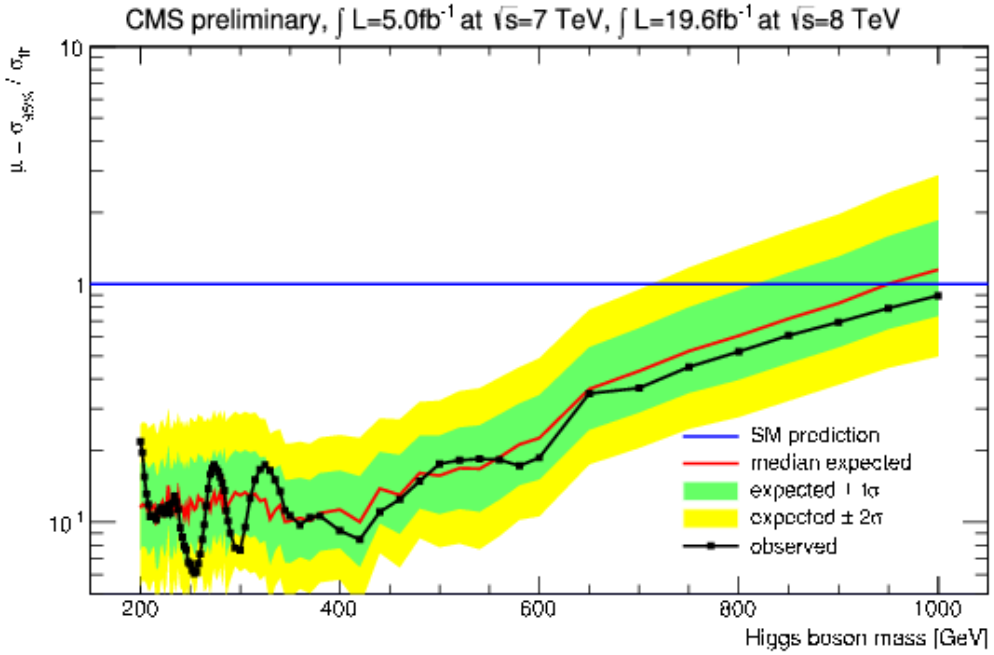


Figure 10.6: Observed (black line) and expected (red line) 95% CL upper limit on the ratio of the product of the production cross section and branching fraction to the SM expectation for the Higgs boson decaying into two Z bosons. The horizontal solid line at unity indicates the SM expectation. Result is combination of all individual channels: $H \to ZZ \to 4l$, $H \to ZZ \to 2l2\nu$ and $H \to ZZ \to 2l2q$.

Chapter 11

Conclusions

The analysis performed on the $H \to ZZ \to l^+l^-q\bar{q}$ final state in the CMS Collaboration have been presented. This particular channel is relevant when the two Z bosons are on-shell, for mass hypothesis over 180 GeV, due to the larger production rate compared to other channels, and due to the possibility to fully reconstruct the invariant mass of the products of the decay. This allows to search for resonances in the m_{ZZ} spectrum and perform a sophisticated shape-based analysis. This channel significantly contributes to the Higgs searches in the range 200−600 GeV, which had not been explored before the LHC.

This thesis has first introduced the theoretical background of the mass mechanism, to continue with the description of the LHC and CMS designs. In chapter 6 the studies performed on the DT chambers with the first data from collisions collected by CMS have been presented. DT chambers are one of the responsibilities of the CIEMAT group in the CMS Collaboration, and a key component of the CMS detector. A mapping of all cells has been done and very high efficiency have been observed throughout the DT muon system, with an average efficiency of 97.7% for single hit detection. Resolution of individual cells has also been measured, although the lack of statistics and the changing initial conditions of the CMS setup produced a result less satisfactory than the actual detector behaviour. The impact of background signals in the muon system have been evaluated. A strong and linear dependency have been observed with the instant luminosity of the LHC, and in particular with the beam cross through the detector. A mapping of the detector and the regions suffering more noise have been achieved, and results have been extrapolated to higher luminosity conditions, like the ones reached during 2012 and that are expected when the LHC starts running again. The conclusion (confirmed in 2012) was that the background rates would be under control and would not imply any trouble in the muon reconstruction and the Physics involved.

Right after, in chapter 8 the details of the $H \to ZZ \to l^+l^-q\bar{q}$ analysis have been presented. Two, similar, although independent analysis, have been performed with the 7 TeV and 8 TeV datasets, respectively. The presence of a very large Z+jets

background, as well as other contributions, like $t\bar{t}$ or SM ZZ and WZ production, imposes an optimized selection of events, based on the differences between the signal and background processes, in an attempt of improving signal over background S/B ratio. This includes kinematical differences, as well as spin-parity correlations, of the objects involved in the decay.

In order to improve the sensitivity, the sample is splitted according to the number of b-tagged jets present in the selected candidates, and treated separately.

The different background processes have been determined differently. The minor diboson contribution is extracted from simulation, whereas the $t\bar{t}$ is ascertained from a control sample of $e^{\pm}\mu^{\mp}$ data. The dominant Z+jets background is determined from a control region in the sidebands of the m_{jj} spectrum and from the MC simulation. All the analysis strategy has been agreed on control samples and simulation, to avoid any bias from the signal data in the conclusions taken.

Results, which included an evaluation of all relevant systematic uncertainties associated to the analysis, have been presented in chapter 10 using a frequentist approach. They include the 5 fb^{-1} of data recorded at $\sqrt{s}$ =7 TeV and the 19.6 fb^{-1} recorded at $\sqrt{s}$ =8 TeV. No excess of events over the background prediction have been observed in the range under study, and upper exclusion limits have been set for the SM Higgs production in the range 275−600 GeV at 95% CL. Although exclusion limits have not been set for other models, no indication of any process except for the considered backgrounds have been observed.

Finally, the next steps over the presented results include the extension of the m_H considered up to 1 TeV. This objective implies considering the case of the two original jets merging into one reconstructed jet in the detector. A dedicated analysis studying the VBF production mode will optimize the sensitivity of this channel. And, ultimately, the extension of the SM interpretation to other models, like Supersymmetry or electroweak singlets, is intended. All these improvements are expected in a near term. The prospects in a longer term would be to analyze the data recorded from 2015 at a higher energy in the center of mass (at least $\sqrt{s}$=13 TeV expected) and with higher luminosity conditions, allowing to improve the mass range under study, the sensitivity of the analysis, and to increase the models susceptible of being tested.

Chapter 12

Appendix A: Data and Monte Carlo Simulated Samples

Table 12.1: Data samples used in the analysis.

Channel	Dataset	Luminosity [pb^{-1}]
$2\mu 2q$	/DoubleMu/Run2012A-13Jul2012-v1/AOD	808
	/DoubleMu/Run2012A-recover-06Aug2012-v1/AOD	82
	/DoubleMu/Run2012B-13Jul2012-v4/AOD	4429
	/DoubleMu/Run2012C-24Aug2012-v1/AOD	495
	/DoubleMu/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/DoubleMu/Run2012C-PromptReco-v2/AOD	6394
	/DoubleMu/Run2012D-PromptReco-v1/AOD	7274
$2e2q$	/DoubleElectron/Run2012A-13Jul2012-v1/AOD	808
	/DoubleElectron/Run2012A-recover-06Aug2012-v1/AOD	82
	/DoubleElectron/Run2012B-13Jul2012-v4/AOD	4429
	/DoubleElectron/Run2012C-24Aug2012-v1/AOD	495
	/DoubleElectron/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/DoubleElectron/Run2012C-PromptReco-v2/AOD	6394
	/DoubleElectron/Run2012D-PromptReco-v1/AOD	7274
$e\mu qq$	/MuEG/Run2012A-13Jul2012-v1/AOD	808
	/MuEG/Run2012A-recover-06Aug2012-v1/AOD	82
	/MuEG/Run2012B-13Jul2012-v4/AOD	4429
	/MuEG/Run2012C-24Aug2012-v1/AOD	495
	/MuEG/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/MuEG/Run2012C-PromptReco-v2/AOD	6394
	/MuEG/Run2012D-PromptReco-v1/AOD	7274

Table 12.2: Background simulated samples of the Summer12 production used in the analysis. The equivalent luminosity of the processed events for each sample is computed using the (N)NLO cross section in the 3^{rd} column.

Process	dataset	σ [pb]	luminosity [fb^{-1}]
Z+jets (inclusive)	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3503.71	8.7
Z+1 jet (exclusive)	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	660.6	36.4
Z+2 jet (exclusive)	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	215.1	101.6
Z+3 jet (exclusive)	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	65.79	167.4
Z+4 jet (exclusive)	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27.59	232.1
$t\bar{t}$	/TTTo2L2Nu2B_8TeV-powheg-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	23.38	461
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	17.654	549
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	22.88	424
WW	/WW_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	57.1097	168

Table 12.3: The signal samples,

$m_H(GeV)$	$\sigma\times$ Br(H $\to$ ZZ $\to$ l$^+$l$^-$q$\bar{\text{q}}$) [pb]
200	0.2566
210	0.2538
220	0.2416
230	0.2278
250	0.2022
275	0.1751
300	0.1563
325	0.1478
350	0.1482
375	0.1360

$m_H(GeV)$	$\sigma\times$ Br(H $\to$ ZZ $\to$ l$^+$l$^-$q$\bar{\text{q}}$) [pb]
400	0.1111
425	0.0914
450	0.7311
475	0.6000
500	0.4719
525	0.0380
550	0.0305
575	0.0250
600	0.0201

Chapter 13

Búsqueda de resonancias masivas similares al bosón de Higgs, en el estado final $\mathrm{H} \to \mathrm{ZZ} \to \mathrm{l}^{+}\mathrm{l}^{-}\mathrm{q}\overline{\mathrm{q}}$ en colisiones pp en el experimento CMS en el LHC

El aparente orden y belleza de la Naturaleza ha motivado a la Humanidad a lo largo de la Historia a preguntarse por las reglas que la rigen. Durante el siglo XX se produjo el mayor avance en la comprensión del Universo, con el desarrollo de la Teoría General de la Relatividad, la Mecánica Cuántica y la Teoría Cuántica de Campos. El Modelo Estándar (SM, en inglés) es la teoría más general y de mayor precisión que se ha podido construir hasta el momento. Describe tres de las cuatro interacciones conocidas entre partículas elementales, y ha sido confirmada por un gran número de observaciones.

Una de las predicciones del SM más difíciles de comprobar experimentalmente ha sido el origen de la masa de las partículas elementales. El mecanismo Brout-Englert-Higgs predice una ruptura espontánea de simetría de un campo escalar adicional. El Gran Colisionador de Hadrones (LHC) en el CERN, y el experimento llamado Solenoide Compacto de Muones (CMS) han sido diseñados y construidos para dar respuesta a gran parte de las cuestiones sin resolver del SM, con especial interés en la búsqueda de un bosón escalar asociado a ese campo, con el que poder confirmar o rechazar dicho mecanismo. El descubrimiento durante 2012 de un bosón escalar con masa m_H∼125 GeV, compatible hasta el momento con la predicción del SM, por los experimentos CMS y ATLAS del LHC en el CERN ha sido uno de los mayores descubrimientos científicos de las últimas décadas. Sin embargo, el sector escalar todavía comienza a comprenderse ahora, y dada la certeza de que el SM no es una teoría completa, muchos modelos adicionales han surgido prediciendo resonancias adicionales con signaturas similares

al bosón de Higgs. La comprobación experimental de estos modelos, como la Supersimetría, es también una de los objetivos del LHC.

Esta tesis describe la búsqueda realizada dentro de la Colaboración CMS de resonancias producidas por una partícula decayendo en dos bosones Z, uno de los cuales se desintegraría leptónicamente, mientras el otro lo haría hadrónicamente: $H \to ZZ \to l^+l^-q\overline{q}$. Para ello se ha utilizado la muestra de datos recogida durante 2010, 2011 y 2012, en colisiones protón-protón producidas en el LHC a una energía de $\sqrt{s}$ =7 TeV y $\sqrt{s}$ =8 TeV en el centro de masa. La luminosidad integrada recogida en el primer caso es de $\mathcal{L}$ =5.0 fb^{-1} y mientras que en el segundo caso es de $\mathcal{L}$ =19.6 fb^{-1}.

El estudio del canal $H \to ZZ \to l^+l^-q\overline{q}$ es muy interesante, ya que combina una tasa de producción relativamente alta comparado con otros estados finales, con la posibilidad de calcular la masa invariante de los cuatro objetos reconstruidos. Los resultados del análisis han sido interpretados bajo la hipótesis de un bosón del Higgs del SM, aunque la presencia o no de algún exceso de datos, o discrepancia con respecto a las prediciones de los procesos de fondo, podría ser una indicación de Nueva Física más allá del Modelo Estándar.

La tesis sigue la siguiente estructura: El capítulo 3 resume el SM y el mecanismo de Brout-Englert-Higgs para explicar la masa de las partículas fundamentales, incluyendo los procesos de producción esperados en el LHC y sus modos de decaimiento. También se incluyen los resultados obtenidos en el momento de escribir la tesis sobre el descubrimiento de un bosón escalar con masa m_H∼125 GeV. El capítulo 4 describe brevemente el diseño del LHC y su funcionamiento durante la primera fase de toma de datos, que duró hasta febrero de 2013. El capítulo 5 a su vez muestra una detallada descripción del detector CMS y todos los subsistemas involucrados en la detección.

La Colaboración CMS debe su nombre al gran esfuerzo empleado en la identificación y reconstrucción de muones con la mayor precisión posible, ya que muchos de los procesos de interés involucran a muones en el estado final, incluyendo al $H \to ZZ \to l^+l^-q\overline{q}$. Por tanto, el detector de muones es una pieza vital de CMS. El capítulo 6 describe el diseño y funcionamiento de las Cámaras de Deriva (DT), responsables de la detección y medida de muones que crucen la parte central del detector. El CIEMAT ha sido responsable del diseño, construcción y mantenimiento de esas cámaras. El capítulo se centra en los resultados obtenidos en la medida de la eficiencia y la resolución de los tubos de deriva, con los primeros datos de colisiones producidos en el LHC. Una buena comprensión del comportamiento de las cámaras es importante en la subsiguiente reconstrucción de muones, y para mejorar la respuesta del detector. Finalmente, se ha analizado la caracterización e influencia del ruido de fondo en las cámaras.

El capítulo 7 describe los objetos involucrados en el estado final $H \to ZZ \to l^+l^-q\overline{q}$. Estos objetos son muones, electrones, chorros hadrónicos y energía faltante. También se reseñan otras magnitudes relativas a las características de los objetos, como aislamiento o identificación de hadrones pesados.

Tras la presentación de todos los ingredientes, en el capítulo 8 se detalla el análisis $H \to ZZ \to l^+l^-q\overline{q}$ desarrollado con la muestra de datos a 8 TeV recogidos durante 2012 y 2013. Se trata del capítulo central de esta tesis. En él se explican los rasgos característicos de este canal. La producción especialmente alta de un bosón Z asociado con chorros hadrónicos procedentes de procesos QCD (lo que se conoce comúnmente como Z+jets) es cinco órdenes de magnitud mayor que la señal esperada, convirtiéndola en el mayor proceso de fondo al que enfrentarse, ya que en muchos casos son indistinguibles. Otros fondos son la producción de pares top-antitop y la producción normal de pares de bosones (ZZ, WZ). En consecuencia, una selección optimizada de eventos es necesaria, con el objeto de maximizar la fracción de sucesos de señal en la muestra. Tras ello, se detalla cómo se han determinado los procesos de fondo y de la señal de Higgs. A continuación se presenta una pormenorizada descripción de las incertidumbres sistemáticas y cómo se han cuantificado. Finalmente, se presentan los resultados obtenidos. Se han realizado dos análisis independientes con las muestras de datos a $\sqrt{s}$ =7 TeV y a $\sqrt{s}$ =8 TeV. El análisis más reciente ha incluido diferentes mejoras, y las condiciones de la toma de datos también variaron. Sin embargo, debido a la gran similitud en las estrategias seguidas en ambos casos, sólo se han enumerado las diferencias entre uno y otro, en el capítulo 9.

Por último, en el capítulo 10 se presenta la interpretación estadística de los restultados, bajo la hipótesis de la existencia de un bosón de Higgs del SM y los límites de exclusión obtenidos. La interpretación bajo la hipótesis del SM sólo debe de tomarse como punto de referencia para la existencia de indicaciones de otros modelos. Y para terminar, el capítulo 11 resume el trabajo expuesto a lo largo de la tesis, las conclusiones obtenidas y las perspectivas para mejorar los resultados obtenidos.

Mi contribución a esta tesis incluye, por un lado, el análisis del comportamiento de las cámaras de deriva, detallado en el capítulo 6, y que fue publicado en [2, 3]. Por el otro lado, contribuí a los dos análisis $H \to ZZ \to l^+l^-q\overline{q}$ de colisiones a 7 y 8 TeV, descritas en los capítulos 9 y 8, respectivamente. En el análisis de datos a 7 TeV contribuí fundamentalmente en la interpretación estadística del análisis y en el cálculo de límites, aunque también participé en la estrategia de selección de sucesos, así como en la validación de las simulaciones computacionales. En el análisis de 8 TeV mi contribución fue mucho más significativa, participando en la mayoría de los aspectos del análisis: en la selección de sucesos, en la determinación de procesos de fondo y de señal, y en el cálculo de límites a la producción del Higgs. Este trabajo ha sido publicado en [4, 5, 6].

A lo largo del texto, se han utilizado unidades naturales h=c=1.

13.1 El mecanismo de masa en la Física de Partículas

El Modelo Estándar describe con gran precisión tres de las cuatro interacciones fundamentales observadas en la Naturaleza entre partículas elementales: electromagnética, nuclear fuerte y nuclear débil. Además, clasifica las partículas de acuerdo con su espín (fermiones y bosones) y de acuerdo con su carga de color (leptones y quarks). Desde el punto de vista formal, el SM es una Teoría Cuántica de Campos basada en los grupos de simetría $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. Sin embargo, este modelo no satisface la observación empírica de que los fermiones, así como los bosones mediadores de la interacción débil, son masivos. El modelo Brout-Englert-Higgs (BEH) resuelve la cuestión mediante la ruptura espontánea de simetría de un campo escalar adicional. En su formulación más sencilla, este campo tendría asociada la existencia de un solo bosón escalar, comúnmente llamado *bosón de Higgs* del SM. Por lo tanto, la búsqueda experimental de dicha partícula sirve para respaldar o rechazar el modelo BEH. En el LHC, el proceso dominante de producción del bosón de Higgs es la fusión de gluones (gg), a través de un bucle de quarks top. Otros procesos que contribuyen a la producción total son la fusión de bosones vectoriales (VBF) y la producción asociada a otras partículas, bien sean quarks top o bosones débiles (ttH,WH,ZH). Por otro lado, los canales de decaimiento son varios y dependen de la masa de la partícula, la cual es un parámetro desconocido. Por lo tanto, es necesario llevar a cabo diferentes análisis para diferentes hipótesis de masa. En el gráfico de la izquierda de la figura 13.1 se muestra la sección eficaz de la producción del bosón de Higgs para una energía de $\sqrt{s}$=8 TeV en el centro de masas en el LHC, en función de la hipotética masa del Higgs. La figura de la derecha, en cambio, muestra la fracción de eventos (BR) asociada a cada proceso de desintegración, en las mismas condiciones.

Aquellos canales que presenten una mayor producción en el LHC, comparado con otros procesos que presenten la misma signatura final (normalmente llamados *fondo*) serán los óptimos para la búsqueda del Higgs. Ésto ha quedado patente con el descubrimiento, en 2012, de un bosón de Higgs compatible con el SM de masa m_H=125.7±0.3 (stat.) ±0.3 (syst.) GeV, ya que los canales más sensibles a la señal fueron el estado final de cuatro leptones $H \to ZZ \to l^+l^-l^+l^-$ y de dos fotones $H \to \gamma\gamma$, cuya producción no es la dominante. Además, tanto CMS como ATLAS han mostrado evidencia del bosón de Higgs decayendo en dos bosones $H \to WW$ y directamente en fermiones $H \to b\bar{b}$ y $H \to \tau\bar{\tau}$.

Sin embargo, a pesar de su éxito predictivo, es bien sabido que el SM no es una teoría completa, ya que es incapaz de explicar varios hechos experimentales, como son: el origen de la Materia Oscura, la asimetría materia-antimateria del Universo, la masa de los neutrinos, la gravedad, o la razón por la que sólo se han observado tres familias de partículas en la Naturaleza.

Para resolver todas estas cuestiones se han planteado diferentes modelos, como

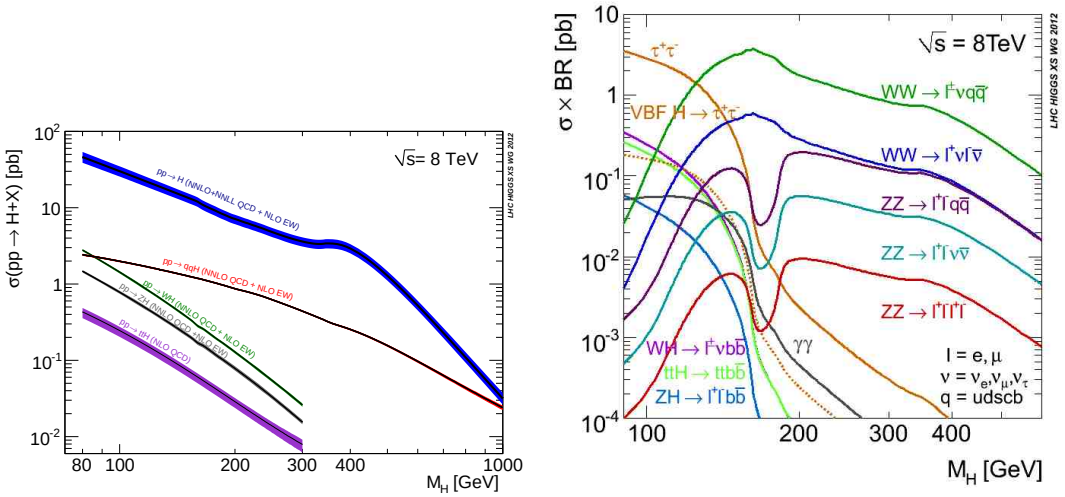


Figure 13.1: Izquierda: sección eficaz de diferentes procesos de producción del bosón de Higgs para una energía de $\sqrt{s}$=8 TeV en el centro de masas en el LHC, en función de la hipotética masa del Higgs. Derecha: Producto de la sección eficaz del bosón de Higgs por la tasa de decaimiento de diferentes canales, para una energía de $\sqrt{s}$= 8 TeV en el centro de masas en el LHC, en función de la hipotética masa del Higgs.

Supersimetría, Teoría de cuerdas, etc. En lo concerniente al mecanismo de masa, todavía es temprano para afirmar que el bosón de Higgs descubierto es exactamente el predicho por el SM, a la luz de las incertidumbres actuales. Por ello, muchas de estas extensiones del SM predicen resonancias adicionales, con la misma signatura del bosón de Higgs. Encontrar alguna de ellas indicaría la existencia de Nueva Física en el rango del TeV, que es el objeto de estudio del LHC.

13.2 El Gran Colisionador de Hadrones

El Gran Colisionador de Hadrones (LHC) es el mayor acelerador de partículas del mundo, ubicado en el CERN. Con forma circular y 26.7 km de largo, acelera las partículas y las agrupa en haces, que circulan en sentidos opuestos. Los haces colisionan en cuatro puntos de la circunferencia, donde se ubican cuatro detectores correspondientes a cuatro experimentos diferentes: ALICE, ATLAS, CMS y LHCb. Mientras que ALICE y LHCb tienen un propósito más específico, ATLAS y CMS son llamados multipropósito, ya que aspiran a estudiar todos los procesos de interés en el rango de 1 TeV.

Dependiendo del tipo de investigaciones que se quieran realizar, por el acelerador pueden circular iones de plomo Pb^{+82} o protones. La mayor parte del programa científico del LHC, incluyendo la búsqueda de bosones de Higgs, se realiza en colisiones protón-protón (pp), por lo que ha ocupado la mayor parte del tiempo. Las partículas circulan por tubos de ultra-vacío, siendo aceleradas en cavidades de radiofrecuencias. Una serie de imanes superconductores controlan la forma del haz y su trayectoria. El

acelerador comenzó a funcionar en noviembre de 2009 y ha ido mejorando sus prestaciones hasta su parada en febrero de 2013. Durante 2010 y 2011 produjo colisiones con una energía de $\sqrt{s}$=7 TeV en el centro de masas. En adelante, la energía fue subida a $\sqrt{s}$=8 TeV. Del mismo modo, la luminosidad instantánea alcanzada ha ido aumentando progresivamente desde los $\mathcal{L} \sim 10^{30} cm^2 \cdot s^{-1}$ hasta $\mathcal{L} \sim 7.7 \cdot 10^{33} cm^2 \cdot s^{-1}$. Desde febrero de 2013, el acelerador se encuentra en una fase de parada técnica con el objetivo de mejorar sus prestaciones, aumentar la luminosidad y la energía por encima de $\sqrt{s}$=13 TeV. Se espera que reanude su actividad en el año 2015.

13.3 El detector CMS

El detector "Compact Muon Solenoid" (CMS) debe su nombre al solenoide superconductor que genera un fuerte campo magnético en su interior, con el objetivo de curvar la trayectoria de las partículas cargadas y medir con precisión su momento lineal transverso al eje de colisión. El detector tiene forma cilíndrica, de 16 m de altura y 21.7 m de largo, dispuesto alrededor del eje del haz. Se divide en una parte central y dos tapas a los lados y le confieren hermeticidad, sumando más de 14.000 toneladas de peso. CMS cuenta con diferentes capas concéntricas de material detector, cada una especialmente diseñado para detectar un tipo de partículas distinto. Del interior al exterior:

- Un detector de trazas de silicio (TRK), dispuesto a pocos centímetros del punto de interacción, que mide con gran precisión la trayectoria de las partículas con carga eléctrica. Ésto le permite, además, calcular su momento lineal con gran precisión.

- Un calorímetro electromagnético (ECAL) hecho de cristales de $WPbO_4$ que permite detectar e identificar electrones y fotones, absorbiéndolos y transformándolos en una señal eléctrica proporcional a la cantidad de energía depositada.

- Un calorímetro hadrónico (HCAL) hecho de capas intercaladas de material centelleante y latón, detecta e identifica cascadas hadrónicas, midiento también su energía.

- Un imán superconductor, solenoidal, que crea un campo constante de B=3.8 T en su interior, donde se sitúan los anteriores detectores. El campo de retorno fluye por una estructura de hierro, la cual además confiere robustez al conjunto.

- Un detector de muones, formado por cámaras de deriva (DT), cámaras de tiras catódicas (CSC) y cámaras de placas resistivas (RPC), todos ellos detectores gaseosos. Las DT y las RPC se localizan en la parte central, mientras que CSC y RPC se encuentran en las tapas laterales. En conjunto, consiguen, por un lado, detectar e identificar muones, y por el otro, contribuir a la medida del momento transverso del muon, especialmente cuando éste es alto.

La gran hermeticidad del conjunto permite estimar la "energía faltante" en cada evento, indicación de la existencia de partículas no detectables por el detector, como los neutrinos. Finalmente, debido a la gran cantidad de colisiones que tienen lugar en el punto de interacción (∼400 millones de colisiones/s), es necesario un sistema de filtrado que seleccione solamente aquellos sucesos potencialmente interesantes, de modo que la muestra recogida sea manejable. El llamado "*trigger*" reduce la tasa de sucesos seleccionados a ∼100/s.

13.4 Funcionamiento de las cámaras de deriva en el sistema de muones

Las DT son una parte fundamental de CMS, encargadas de detectar el paso de muones por la zona central del detector, y reconstruir su trayectoria a su paso por el sistema de muones. Hay en total 250 cámaras, dispuestas en cuatro niveles (o *estaciones"*) concéntricos. Cada cámara está formada por sucesivas capas paralelas al eje del haz, y cada capa está formada a su vez por una serie de celdas gaseosas de deriva. En la figura derecha de 13.2 se muestra un esquema de la estructura de una cámara. Cuando un muon atraviesa una celda, genera una corriente de deriva, que es recogida en el ánodo y transformada en un tiempo y una posición del paso de la partícula. El ajuste de todas las señales recibidas en una cámara permite reconstruir un segmento lineal que contiene información de la posición y la dirección de la trayectoria de la partícula. El CIEMAT ha sido responsable del diseño, construcción y mantenimiento de las DT. La correcta caracterización de las DT ha sido fundamental al principio de la toma de datos para la comprensión y mejora de los datos reconstruidos por el detector. Para ello, se ha estudiado diferentes magnitudes.

La eficiencia de reconstrucción mide la probabilidad de obtener una señal cuando un muon atraviese una determinada capa. Se han distinguido dos casos: cuando las señales medidas ("hits") están necesariamente asociadas a una traza reconstruida, y cuando no es necesario. En el primer caso, la eficiencia promedio es del 96.1% en todas las capas, y en el segundo caso del 97.7%, con un comportamiento bastante homogéneo en todo el detector. Es necesario remarcar que el diseño de CMS en general, y de las cámaras de deriva en particular, es altamente redundante, por lo que cualquier ineficiencia de este tipo en ningún caso implicaría la no detección de un muón, ni siquiera un empobrecimiento en la medida de su momento transverso. El gráfico de la derecha en la figura 13.2 muestra la eficiencia de una capa en función la la posición de la traza dentro de la capa. La mayor fuente de ineficiencia se debe a muones que atraviesan la capa por la unión entre dos celdas, por lo que la estructura de celdas es visible en la figura.

La resolución espacial de una celda ha sido medida también, pero debido a la baja estadística acumulada y a las condiciones iniciales pre-óptimas de calibración, alineamiento y sincronización de las cámaras, los resultados no consiguieron reflejar correctamente el comportamiento de las celdas de deriva, que se sitúa en σ ∼300 μm.

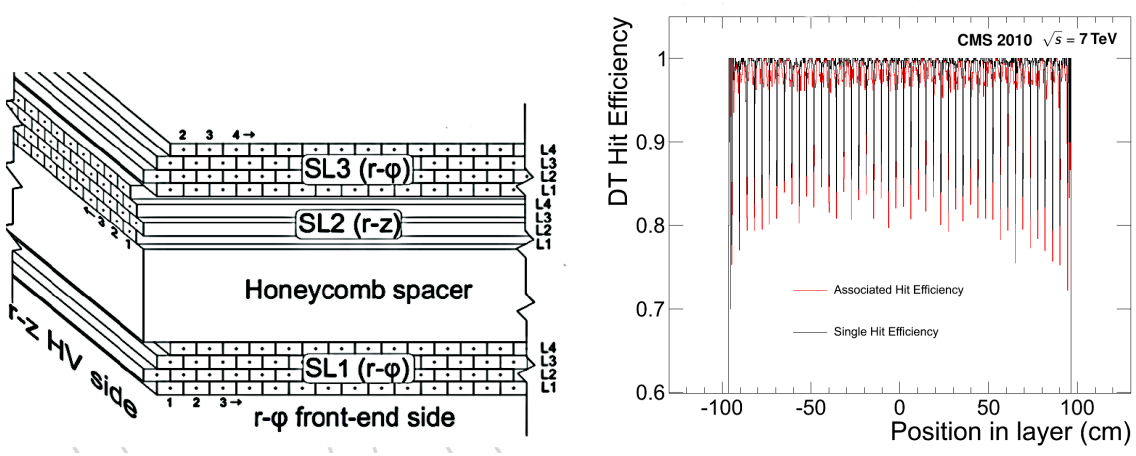


Figure 13.2: Izquierda: Esquema de una cámara de deriva. Derecha: Eficiencia de detección de una capa de una cámara en función de la posición de la traza. La línea negra muestra la probabilidad de encontrar una señal en ese punto, mientras que la línea roja implica que además las señales se han asociado a una traza reconstruida.

Finalmente, se ha caracterizado la cantidad de señales de ruido que reciben las cámaras y su disposición espacial en el detector, así como inferir sus causas. Se ha visto que, además de un número insignificante de celdas que muestran un comportamiento anormalmente ruidoso, el cruce de haces por el centro del detector es la principal fuente de ruido de fondo. Por un lado, induce la existencia de radiación (neutrones principalmente) en la caverna, que rebota en las paredes y afecta principalmente a las cámaras superiores de la estación más externa. Por otro, las cámaras más internas también están sometidas a mayores niveles de radiación. Sin embargo, tanto estos niveles como los estimados para condiciones de alta luminosidad, son relativamente bajos y no suponen ningún problema para el buen funcionamiento de las cámaras ni para la correcta reconstrucción de trazas.

13.5 Reconstrucción de leptones y chorros hadrónicos en el experimento CMS

Este capítulo se centra en la reconstrucción de las partículas estudiadas en el estado final $H \to ZZ \to l^+l^-q\bar{q}$, que son muones, electrones y chorros hadrónicos (comúnmente llamados "*jets*". La reconstrucción de los objetos de un evento se lleva a cabo con el algoritmo "*particle-flow*" (PF), que combina la información de todos los subdetectores de forma global para identificar y reconstruir individualmente cada una de las partículas creadas en la colisión, clasificándolas en electrones, fotones, hadrones cargados, hadrones neutros y muones. Un algoritmo como éste mejora la identificación de las partículas y la resolución energética de jets y energía faltante.

La reconstrucción de tanto muones como electrones es muy robusta y eficiente, fruto de la cantidad de capas de detección redundantes. Se siguen diferentes criterios de calidad para garantizar la pureza de los leptones reconstruidos. Criterios de calidad muy estrictos pueden afectar a la eficiencia de identificación, por lo que se establecen varios puntos de trabajo con diferente relación eficiencia-pureza, en función de las necesidades de cada análisis de CMS. Un parámetro importante de los leptones analizados es el *aislamiento*, que mide la cantidad de energía/momento recogida en las inmediaciones del leptón, y permite distinguir leptones procedentes de un jet o del decaimiento de un bosón débil. Sin embargo, el aumento de la luminosidad instantánea implica un aumento de colisiones simultáneas, y en consecuencia, de radiación dentro del detector, lo que puede sesgar el valor del aislamiento. Para ello se han dispuesto diferentes técnicas de corrección de dicho parámetro.

Los resultados arrojan una eficiencia promedio superior al 90%. La resolución en energía de los electrones se sitúa entre el 1-4%, mientras que la resolución en la medida del momento transverso de los muones se encuentra entre el 1-10% para muones de muy alto momento (por encima de 1 TeV).

La construcción de jets se lleva a cabo a partir de las partículas obtenidas con la técnica PF. El algoritmo "*anti-kt*" asocia las partículas y forma un jet cónico, en función de sus momentos transversos y las distancias entre ellas. Una vez el jet ha sido formado, se llevan a cabo diferentes correcciones, fruto de la calibración, para adecuar la respuesta en energía obtenida en el detector a la energía verdadera del jet. Así, finalmente se obtiene una resolución en energía superior al 10% en jets de menos de 100 GeV.

La identificación de jets procedentes de desintegraciones de quarks b es muy importante en muchos procesos de interés, como física del quark top, o búsquedas de Higgs como la objeto de esta tesis. Se han desarrollado diferentes técnicas para identificar un jet como procedente de un quark b, que explotan su larga vida media, su alta masa o su fragmentación. La eficiencia de identificar correctamente un jet "b" varía inversamente con la probabilidad de asociar erróneamente un jet procedente de un quark ligero como procedente de un quark b, conocido en inglés como *mistags*.

Aunque el detector es altamente hermético, la suma de momentos puede no estar balanceada, debido a la presencia de neutrinos, que son indetectables por CMS, a la propia aceptancia, o a la resolución del aparato. La energía faltante ($\not{E}_{\mathrm{T}}$) se define como el opuesto de la suma de momentos de todas las partículas. Otra variable muy utilizada para discriminar si la energía faltante se debe a efectos de resolución o a la existencia de neutrinos en el estado final es la "significancia de la energía faltante", definida como $\lambda = 2ln\left(\frac{\mathcal{L}(\vec{\varepsilon}=\vec{\varepsilon_T})}{\mathcal{L}(\vec{\varepsilon}=0)}\right)$

13.6 Búsqueda del bosón de Higgs en el estado final $H \to ZZ \to l^+l^-q\bar{q}$ en colisiones pp a $\sqrt{s}$ =8 TeV

La búsqueda de bosones de Higgs con una masa $m_H \sim 2m_Z$ son muy interesantes porque no habían sido explorados antes del LHC y su descubrimiento tendría implicaciones en la existencia de Nueva Física más allá del SM. El estado final $H \to ZZ \to l^+l^-q\bar{q}$ es muy atractivo porque permite reconstruir la masa invariante de los cuatro objetos, pudiendo así estudiar el espectro m_{ZZ} en busca de resonancias o excesos. Además, debido a que el bosón Z decae principalmente en quarks, este estado final presenta un equilibrio entre la tasa de producción (mucho más alta que en otros estados finales sin quarks) y una cantidad asumible de procesos de fondo en el LHC que esconden la señal (lo que no sucede en estados finales con cuatro quarks).

Incluso después del descubrimiento de un bosón de Higgs con una masa $m_H \sim$125 GeV, sigue siendo un canal muy interesante, ya que muchos modelos predicen la existencia de resonancias adicionales a más alta masa, y las incertidumbres actuales asociadas a las propiedades del bosón de Higgs de 125 GeV no excluyen su existencia. En todos estos modelos, el decaimiento a bosones WW y ZZ todavía domina, por lo que este canal puede aportar mucha información al respecto.
Este análisis busca dos leptones del mismo sabor y carga opuesta (electrones o muones) resonantes con la masa del bosón Z, y dos chorros hadrónicos también resonantes con la masa del bosón Z.

Desafortunadamente, otros procesos producidos en el LHC dejan la misma señal en el detector, por lo que es necesaria una estrategia de filtrado para obtener una muestra lo menos contaminada posible. El principal proceso de fondo es la producción de un bosón Z asociado con la producción de dos jets fruto de procesos de QCD (comúnmente llamado *Z+jets*). La producción de este tipo de sucesos, muchas veces indistinguibles del Higgs, es 10^5 veces superior a la esperada para el Higgs. Para optimizar la sensitividad del análisis, la muestra se divide en 6 canales exclusivos. Por un lado se separan los sucesos con electrones y muones, y por otro se divide la muestra según el número de jets identificados como provenientes de un quark b: 0, 1 o 2 jets en el suceso. Así se aprovecha que el estado final $H \to ZZ \to llb\bar{b}$ presenta una razón señal/fondo (S/B) mucho más alta, ya que la presencia de jets "b" es mayor en el decaimiento del Higgs que en la producción de Z+jets del SM en el LHC. Esta categoría tiene una estadística bastante baja, que se ve compensada con la categoría con cero quarks b, que a pesar de tener peor S/B presenta una alta estadística. Combinando los seis canales se consigue optimizar la sensitividad.

Precisamente es en la categoría $llb\bar{b}$ donde más importancia tiene el siguiente proceso de fondo: la producción de pares top-antitop ($t\bar{t}$) que en su decaimiento a leptones produce una señal confundible.

En ambos casos, a través de la topología del evento, y la cinemática de los objetos reconstruidos, es posible filtrar en buena medida muchos de estos procesos de fondo y obtener una medida sensible de la hipotética producción del bosón de Higgs.

Finalmente, también se considera la producción en el SM de pares de bosones ZZ, WZ, WW, que pueden tener una signatura idéntica al Higgs. Por suerte, la producción de estos procesos es muy pequeña comparado con los anteriores.

Toda la estrategia del análisis se ha establecido estudiando simulaciones Monte-Carlo y datos en regiones de control, sin mirar las regiones donde uno espera señal de Higgs. La razón es evitar cualquier sesgo en la optimización del análisis o en la determinación de los procesos de fondo. Sólo una vez que la estrategia ha sido definida, se ha procedido a analizar la muestra de la región de señal de Higgs.

Se han llevado a cabo dos análisis independientes, aunque bastante similares, con los datos recogidos durante 2010 y 2011 de colisiones a $\sqrt{s}$=7 TeV, sumando una luminosidad integrada de $\mathcal{L}$=5.0 fb^{-1}, y durante 2012 y 2013 de colisiones a $\sqrt{s}$=8 TeV, habiendo recogido $\mathcal{L}$=19.6 fb^{-1}. Dada la similitud entre ambos análisis, esta tesis se centra en el análisis realizado con datos a $\sqrt{s}$=8 TeV, y tan sólo se enumeran las diferencias existentes entre ambos análisis.

Además de los datos recogidos por CMS, se han utilizado muestras de simulación Monte-Carlo de todos los procesos de señal y fondo considerados en el análisis.

Se han seleccionado solamente sucesos reales que hayan disparado un trigger que exija dos muones o dos electrones con momentos transversos mayores que p_T>17 y p_T>8 GeV, respectivamente. El número de sucesos simulados se ha escalado apropiadamente acorde a la eficiencia de que un determinado suceso con dos leptones efectivamente active el trigger.

13.6.1 Selección de sucesos

La selección de sucesos incluye una primera parte, llamada *preselección*, en la que se exigen ciertos criterios de calidad a los diferentes objetos involucrados: muones, leptones y chorros hadrónicos. Se trata de criterios de calidad en la reconstrucción de las partículas y cortes cinemáticos en su momento transverso y pseudorapidez. Además se exige que los leptones estén aislados. El aumento de interacciones simultáneas es tenida en cuenta a la hora de corregir los criterios de aislamiento y a la hora de seleccionar los jets. Por ejemplo, la variable β discrimina jets cuyos componentes proceden mayormente del vértice primario y los que tienes contribuciones importantes de partículas producidas en otras interacciones.

Para filtrar aquellos pares de partículas que no procedan de un bosón Z, se requiere que la masa invariante de los dos leptones se encuentre en la ventana [76,106] GeV. En la Fig. 13.3 se puede ver la distribución de la masa invariante de leptones, para ambos casos. Los puntos representan los datos medidos, mientras que los histogramas representan la estimación de los procesos de fondo. Del mismo modo, se pide que la masa invariante de los dos jets se encuentre en la ventana [71,111] GeV. En la Fig. 13.4 se muestra la masa invariante de los dos jets después de la preselección. Debido a la peor resolución energética de los jets, la resonancia del Z es más ancha.

Tras la preselección, varias variables definen la selección final. Las partículas producto del decaimiento de un bosón de Higgs presentan correlaciones angulares,

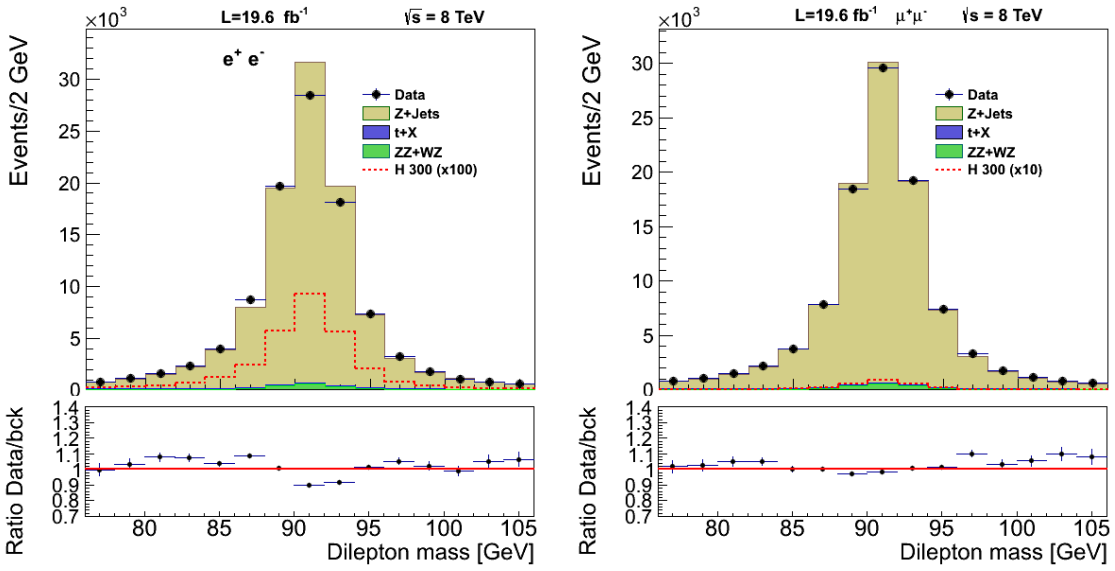


Figure 13.3: Masa invariante del sistema dileptónico, representado para e^+e^- (izquierda) y $\mu^+\mu^-$ (derecha). Los puntos representan los datos, el histograma ocre la estimación de Z+jets, el verde la simulación de dibosones y el azul oscuro la estimación de fondo de top.

impuestas por el espín y la paridad del Higgs. Estas correlaciones no aparecen en el fondo de Z+jets, por lo que se puede construir un distriminante multivariado que mejore la separación señal-fondo.

El gráfico de la izquierda en la Fig. 13.5 muestra los cinco ángulos que describen completamente la cinemática del decaimiento del Higgs, mientras que el gráfico de la derecha muestra la distribución del discriminante angular donde se observa el poder de separación del fondo y la señal.

Los procesos de $t\bar{t}$ producen dos neutrinos en su estado final, por lo que un corte en la energía faltante transversa eliminará gran parte de esta contribución. En nuestro caso, se ha preferido establecer un corte en la significancia de la $\not{E}_{\mathrm{T}}$<10.

La tabla 13.1 resume todos los cortes establecidos en la selección de sucesos.

13.6.2 Determinación de los procesos de fondo y señal.

Una vez establecida la muestra final, se han determinado los diferentes procesos a considerar. Para el proceso Z+ jets se han utilizado muestras de simulación y la región de control definida a los lados del pico en la distribución m_{jj}, donde apenas se espera señal, aunque la topología del suceso se espera similar. Debido a que la simulación presentaba pequeñas diferencias en el espectro $\mathrm{p}_{\mathrm{T}}^{llqq}$ del momento transverso de los cuatro objetos reconstruidos respecto de las distribuciones de datos, se ha repesado la simulación para corregir esta diferencia. Además, para no depender de la precisión de la simulación de la fracción de sucesos con jets b, la normalización de cada una de

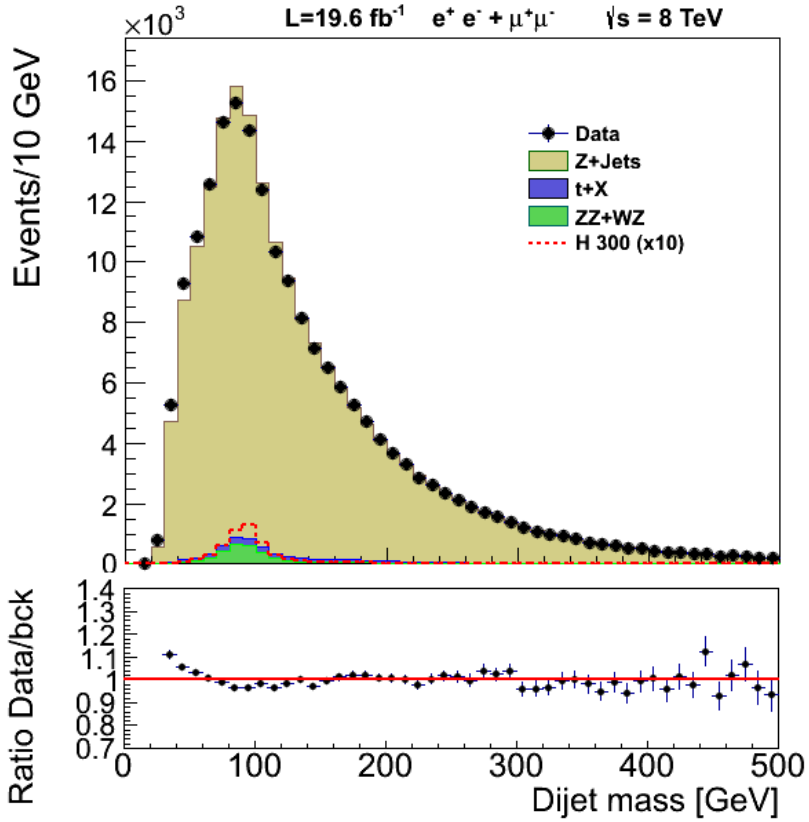


Figure 13.4: Distribución de la masa invariante m_{jj}. Los puntos representan los datos, el histograma ocre la estimación de Z+jets, el verde la simulación de dibosones y el azul oscuro la estimación de fondo de top. La región de señal se define en el rango del espectro de m_{jj} $71<m_{jj}<111$ GeV y la región de control se define a los lados del pico: $60<m_{jj}<71 \cup 111<m_{jj}<130$ GeV.

las tres categorías se ha constreñido a partir del número de sucesos en las región de control, en datos y simulación.

El proceso $t\bar{t}$ se ha extraído de una muestra de datos $e^{\pm}\mu^{\mp}$ con un electrón, un muon y dos jets b. Esta muestra presenta la misma forma que la esperada para el $t\bar{t}$, y la misma cantidad de sucesos que la esperada para los procesos $ee + \mu\mu$ en la simulación. Esta muestra además incluye otros procesos de fondo con una contribución muy pequeña como $t + X$ o el WW. Además, tiene la ventaja de que, al ser una estimación a partir de datos, no se depende de la simulación y se evita el cálculo de ciertas incertidumbres sistemáticas.

Por último, la contribución de dibosones ZZ y WZ se ha estimado directamente a partir de la predicción de la simulación, dada su pequeña tasa de producción respecto de los procesos anteriores, y la fidelidad de la simulación.

La señal se ha determinado también a partir de la simulación. Se ha tenido en cuenta que para masas altas (a partir de $m_H \sim 400$ GeV) una distribución de Breit-Wigner ya

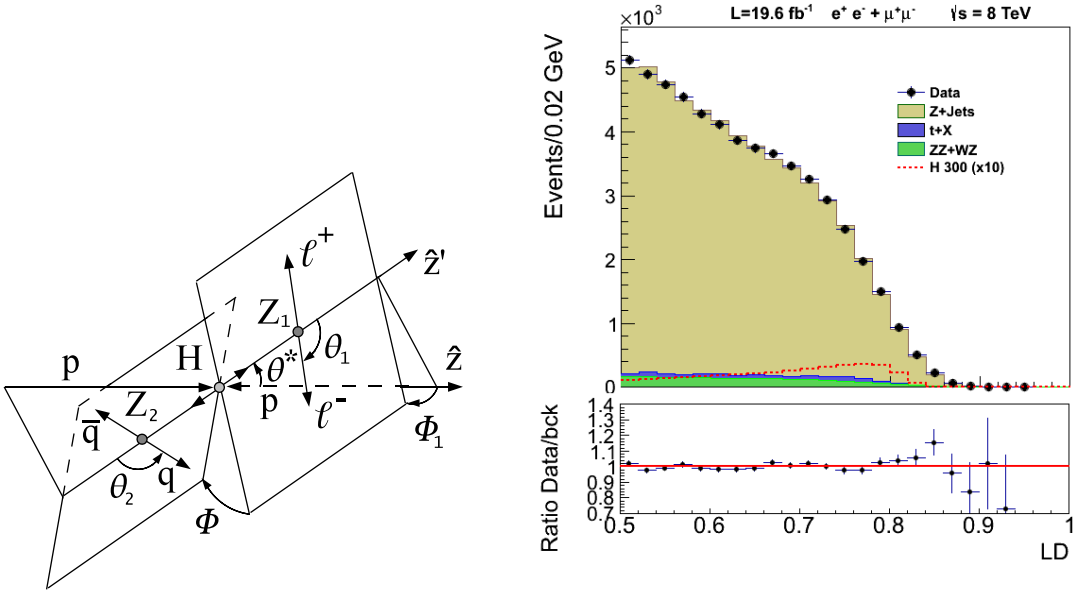


Figure 13.5: Izquierda: esquema del decaimiento de un bosón de Higgs y los cinco ángulos con los que se puede determinar la cinemática del decaimiento. Derecha: distribución del discriminante construido a partir de dichos ángulos, para los datos y las predicciones de fondo y señal.

observable	0 jet b	1 jet b	2 jets b
Preselección			
Calidad del muon	Diferentes variables de la reconstrucción		
Calidad del electron	Diferentes variables de la reconstrucción		
Calidad del jet			
mismo sabor leptónico, carga opuesta			
$p_T(l^{\pm})$	$> 40/20\ GeV$		
p_T(jets)	$> 30\ GeV$		
$\|\eta\|(\ell^{\pm})$	$(e^{\pm}) < 2.5$, $(\mu^{\pm}) < 2.4$		
$\|\eta\|$(jets)	< 2.4		
β del jet	> 0.2		
ΔR	> 0.5		
Selección final			
Identificación jets "b"	ninguno	JPL	JPM & JPL
$m_{\ell\ell}$	[76,106] GeV		
m_{jj}	[71,111] GeV		
helicidad LD	> 0.5		
Significancia de la $\not{E}_T$	< 10		

Table 13.1: Resumen de los requisitos en la selección de eventos.

no reproduce correctamente la forma de la distribución de la masa invariante, sino que es una resonancia mucho más ancha. Para ello se ha utilizado el enfoque "CPS". También se ha tenido en cuenta la interferencia con el proceso $gg \to ZZ$. A medida que aumenta la hipótesis de masa del Higgs, los objetos finales tienen mayor energía cinética, lo que reduce la separación angular entre los productos de cada Z. A partir de m_H∼600 GeV, los dos jets pueden solapar y el algoritmo de reconstrucción puede confundirlos con un único jet. Por lo tanto, la eficiencia en la detección de la señal comienza a disminuir rápidamente. Por esa razón el análisis sólo se ha extendido hasta m_H∼600 GeV. Para estudiar hipótesis más masivas, es necesario un enfoque donde se consideren sucesos con un único jet reconstruido, y establecer alguna técnica para determinar si procede de dos sub-jets. Este enfoque ya ha comenzado a desarrollarse tras la finalización del estudio presentado aquí.

Por último, se han considerado las diferentes fuentes de incertidumbre sistemática que afectan al análisis. Las tablas 13.2 y 13.3 sintetizan todas las incertidumbres consideradas, tanto para los procesos de señal como de fondo.

Causa	0 jets-b	1 jet-b	2 b-tag
Trigger de muones & ID	2.7%		
Trigger de electrones& ID	2%		
Calibración energía electrones	0.2%		
Calibración momento muones	0.1%		
Reconstrucción de jets	1-4%		
b-tagging eff. y tasa mistags	1-4%	1-5%	5-8%
$\not{E}_T$	$< 1\%$		
Núm. interacciones/cruce	1-2%		
Mecanismos de producción (PDF)	1.5%		
Mecanismos de producción (forma)	0-3%		
Luminosidad	4.4%		
Sección eficaz del Higgs (parar R)	13-15%		

Table 13.2: Resumen de incertidumbres sistemáticas en la normalización de la señal. La mayoría de las causas son errores multiplicativos en la medida de la sección eficaz, excepto la incertidumbre en la sección eficaz del Higgs, (que es relevante sólo en la medida de la razón R con respecto al SM).

Finalmente, la distribución de la masa invariante de los cuatro objetos se presenta en la Fig. 13.6 para las seis categorías (2 leptones⊗ 3 categorías de jets b). Se han comparado los datos (puntos) con las estimaciones de cada proceso de fondo (histogramas).

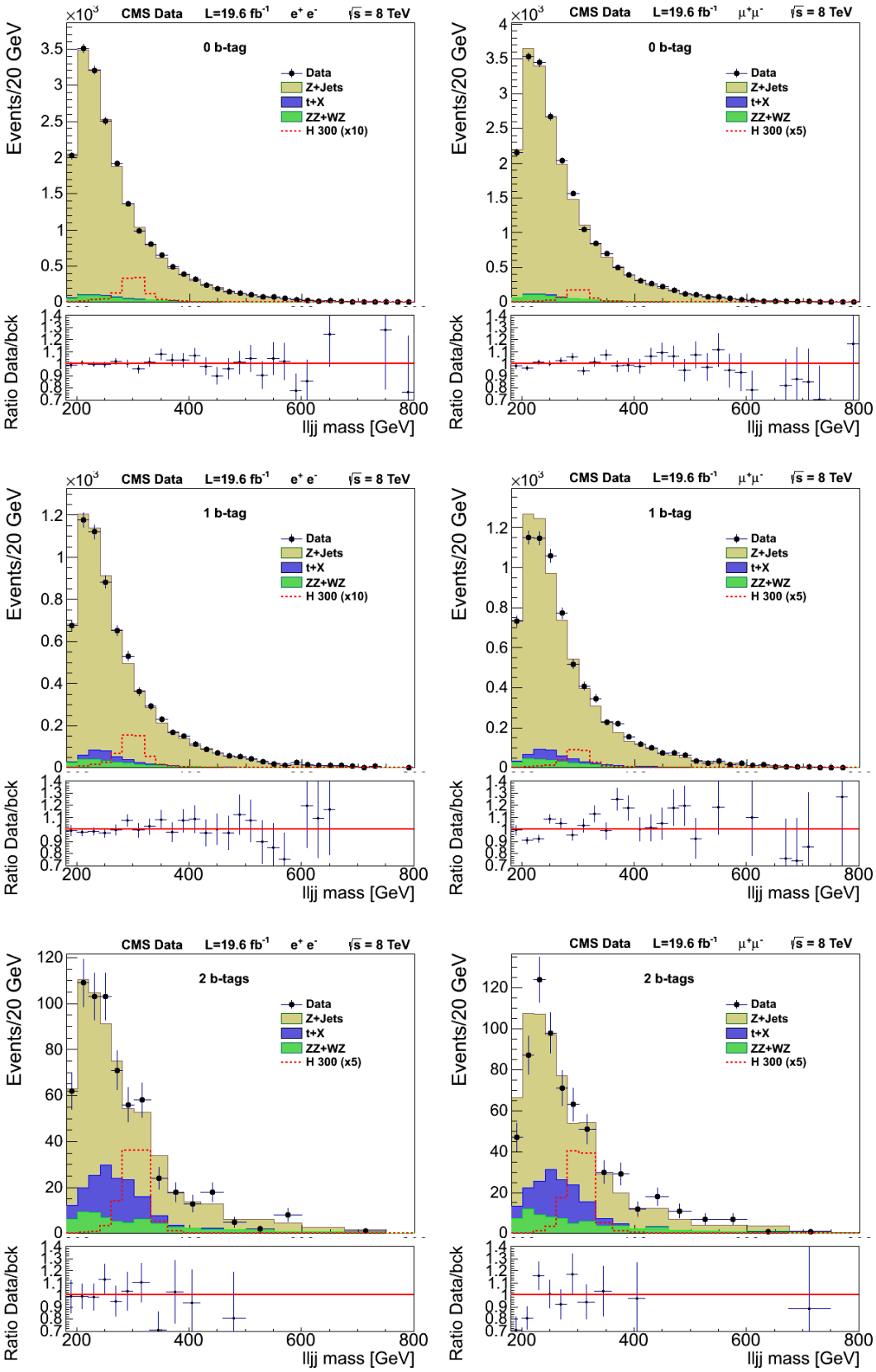


Figure 13.6: Distribución de masa invariante del sistema $lljj$ en la región de señal para en el canal de muones (derecha) y electrones (izquierda). De arriba a abajo, las imágenes corresponden a las categorías de 0,1 y 2 jets-"b". Las incertidumbres sistemáticas no aparecen reflejadas.

Causa	Normalización	Forma
Trigger de muones & ID	2.7%	
Calibración p_T muon	0.1%	
Trigger de electrones & ID	2.0%	
Calibración energía electrones	0.5%	
Calibración energía jets	5.5%	0-4%
Eficiencia identificación jet-b 0-b	+0.4%	
Eficiencia identificación jet-b 1-b	-0.8%	
Eficiencia identificación jet-b 2-b	-4.5%	
Mistags jets-b 0-b	-1.9%	
Mistags jets-b 1-b	+7.8%	
Mistags jet -b 2-b	+6.2%	
$\not{E}_T$	0.3%	
Número de interacciones	0.1%	
Repesado ${p_T}^{\ell\ell jj}$	0.8%	0-3%
Sección eficaz dibosones	15%	
Luminosidad	4.4%	
Diferencia residual		0-15% (0-b)
datos-simulación		0-30% (1-b)
en la región de control		0-40% (2-b)

Table 13.3: Resumen de incertidumbres sistemáticas en la determinación de los procesos de fondo.

13.7 Búsqueda del bosón de Higgs en el estado final $H \to ZZ \to l^+l^-q\bar{q}$ en colisiones pp a $\sqrt{s}$ =7 TeV

El primer análisis de CMS del proceso $H \to ZZ \to l^+l^-q\bar{q}$ fue publicado en 2011 con $\mathcal{L}$=5.1 fb^{-1} recogidos de datos a $\sqrt{s}$=7 TeV. La estrategia del análisis ha sido muy similar a la del análisis a $\sqrt{s}$=8 TeV, aunque existen algunas diferencias debido a las mejoras en la simulación, en la respuesta del detector y en las técnicas de análisis. Asimismo, también cambiaron las condiciones de la toma de datos (energía, número de interacciones simultáneas). Este capítulo resume las principales diferencias llevadas a cabo debido a estos factores.

- Por un lado, en el análisis de 7 TeV, debido a que existían menos interacciones simultáneas, el aislamiento o la reconstrucción de jets eran ligeramente diferentes.
- El discriminante angular y la significancia de $\not{E}_T$ estaban optimizados para cada categoría. En el análisis de 2012 ésto no ofrecía ninguna mejora en el análisis.
- Para intentar mejorar la resolución de la masa invariante (que viene fundamentalmente determinada por la resolución energética de los jets) se llevó a cabo un ajuste cinemático para constreñir la masa invariante m_{jj} hacia el valor nominal de la masa del Z, y modificar m_{llqq} apropiadamente.

- Se utilizaba un discriminante multivariado que utilizaba ciertas propiedades de los jets para intentar discriminar aquellos jets que procediesen de la fragmentación de un quark o de la radiación de un gluon.

- El proceso $t\bar{t}$ estaba determinado completamente a partir de la simulación ya que el análisis a partir de datos $e^{\pm}\mu^{\mp}$ no estaba listo.

- La simulación Z+jets estaba corregida por las diferencias existentes en la masa invariante m_{llqq} en vez de en el momento transverso p_T^{llqq}.

- El discriminante angular y la significancia de $\not{E}_T$ estaban optimizados para cada categoría. En el análisis de 2012 ésto no ofrecía ninguna mejora en el análisis.

- Para intentar mejorar la resolución de la masa invariante (que viene fundamentalmente determinada por la resolución energética de los jets) se llevaba a cabo un ajuste cinemático para constreñir la masa invariante m_{jj} hacia el valor nominal de la masa del Z, y modificar m_{llqq} apropiadamente.

Además, a la hora de realizar el test de contraste de hipótesis acerca de la existencia del bosón de Higgs, en el análisis de 7 Tev se parametrizaron las distribuciones a diferentes funciones, mientras que en el análisis de 8 TeV se han utilizado las formas extraídas directamente de los histogramas, sin parametrizar.

13.8 Interpretación estadística de los resultados

La inferencia estadística permite establecer afirmaciones acerca de la probabilidad de que una medida sea compatible o no con una cierta hipótesis. En esta tésis se ha utilizado la masa invariante m_{llqq} como discriminante para estudiar si las distribuciones obtenidas eran compatibles con la existencia de un bosón de Higgs de masa m_H, dentro del rango $230<m_H<600$ GeV. Al no haberse encontrado ningún exceso de datos en ese rango, por encima de las predicciones del fondo, se han establecido límites de exclusión a la existencia de dicha partícula.

Los límites de exclusión se han establecido a partir de un cierto estadístico q, construido a partir de los datos medidos, de las distribuciones predichas para fondo y señal, y de todas las incertidumbres sistemáticas consideradas, usando el método de máxima verosimilitud, y comparando finalmente los datos con la hipótesis "sólo procesos de fondo" contra "datos de fondo+ señal de Higgs".

La figura 13.7 muestra los límites de exclusión, obtenidos al 95% de nivel de confianza (CL), del cociente de la sección eficaz obtenida con respecto a lo esperado para la existencia de un bosón de Higgs del SM. El gráfico de la izquierda muestra el resultado obtenido en el análisis de datos a 7 TeV, mientras el gráfico de la derecha presenta el resultado obtenidos con datos a 8 TeV. Debido a la mayor cantidad de datos utilizados en éste último, así como a las mejoras realizadas en el análisis, su sensitividad es mucho mayor.

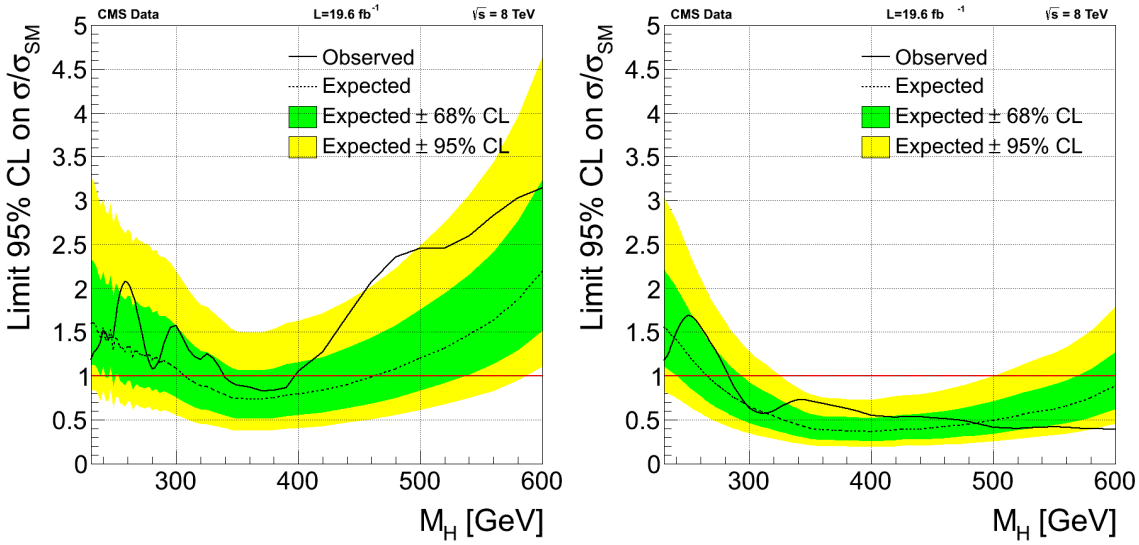


Figure 13.7: Límites superiores observados (línea sólida) y esperados (punteada) al 95% CL en la razón de la sección eficaz producida, con respecto a lo esperado por el SM para el bosón de Higgs. La gráfica de la izquierda presenta los resultados obtenidos en el análisis de datos a 7 TeV, mientras el gráfico de la derecha presenta los resultados obtenidos con datos a 8 TeV. Ambos han sido obtenidos con una técnica CLs frequentista. Los rangos de expectación al 68% y al 95%, para el modelo de "solo fondo" también se muestran con bandas verdes y amarillas, respectivamente. La línea horizontal en la unidad indica la expectación para para un bosón tipo-Higgs.

Ambos resultados han sido combinados, obteniendo unos límites de exclusión final para todos los datos recogidos por CMS que se pueden observar en el gráfico de la izquierda de la Fig. 13.8. La conclusión es que la existencia de un bosón de Higgs como el del SM ha sido excluida al 95% CL en el rango 275-600 GeV. Este estado final también ha sido incluido en la combinación total de todos los canales de de decaimiento $H \to ZZ$. Esta combinación es presentada en el gráfico de la derecha en la Fig. 13.8.

En conclusión, no se ha hallado indicación de la existencia de ningún proceso físico más allá de las predicciones del SM para procesos de fondo. La existencia de un bosón de Higgs con una producción como la predicha por el SM ha sido rechazada al 95% CL en casi todo el rango de estudio. Finalmente, aunque no se han cuantizado límites de exclusión para otro tipo de modelos, no se ha observado nada que pudiese sugerir tal existencia, dentro de la sensitividad de este análisis.

13.9 Conclusiones

En esta tesis se ha presentado el análisis realizado en estado final $H \to ZZ \to l^+l^-q\bar{q}$ en la colaboración CMS. Este canal particular es especialmente relevante cuando los dos bosones Z son reales, lo que sucede para hipótesis de masa de Higgs por encima de 180 GeV, debido a su tasa de producción relativamente alta comparada con otros canales, y debido a que la masa invariante se puede reconstruir a partir de los produc-

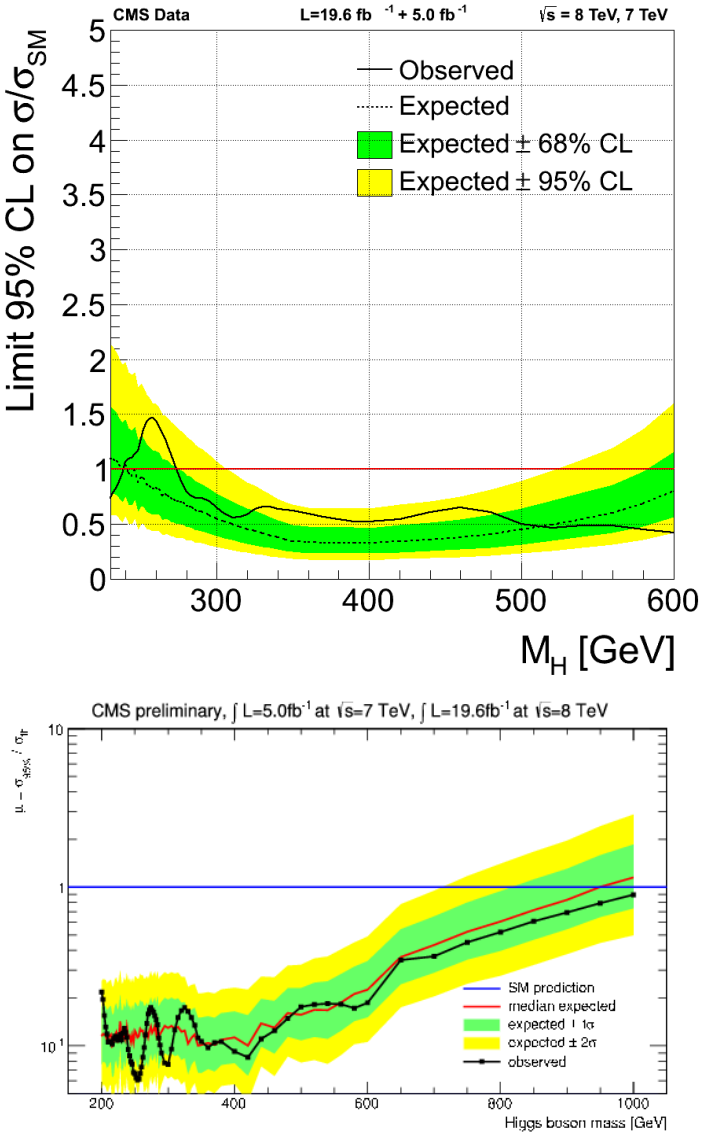


Figure 13.8: Límites superiores observados (línea sólida) y esperados (punteada) al 95% CL en la razón de la sección eficaz producida, con respecto a lo esperado por el SM para el bosón de Higgs. El gráfico superior presenta la combinación de resultados a 7 y 8 TeV del canal $H \to ZZ \to l^+l^-q\bar{q}$, mientras que el gráfico inferior presenta la combinación final de todos decaimientos de $H \to ZZ$: $H \to ZZ \to 4l$, $H \to ZZ \to 2l2\nu$ y $H \to ZZ \to 2l2q$. Los rangos de expectación al 68% y al 95%, para el modelo de "solo fondo" también se muestran con bandas verdes y amarillas, respectivamente. La línea horizontal en la unidad indica la expectación para para un bosón tipo-Higgs.

tos del decaimiento. Este hecho permite buscar resonancias en el espectro de m_{ZZ} y realizar un análisis sofisticado basado en la forma de dicha distribución. Este canal contribuye significativamente a las búsquedas de Higgs realizadas en el rango de 200−600 GeV, que no había sido explorado antes del LHC.
En primer lugar, esta tesis ha introducido el bagaje teórico del mecanismo de masa, para continuar con la descripción de los diseños del LHC y de CMS. En el capítulo 6 se han presentado los estudios desarrollados en las cámaras de deriva con los primeros datos de colisiones del LHC recogidos. Las DT son una de las responsabilidades del grupo del CIEMAT en la colaboración CMS, y una parte vital del detector. Se han analizado todas las celdas de deriva, y se ha observado una eficiencia muy alta en todo el conjunto, con una eficiencia de señales individuales promedio de 97.7%. La resolución espacial de las celdas individuales también se ha medido, aunque la falta de estadística, unido a las cambiantes condiciones iniciales de la configuración de CMS indujeron un resultado menos satisfactorio que el comportamiento real del detector. También se ha evaluado el impacto del ruido de fondo en el sistema de muones. Se ha observado una dependencia fuerte y lineal con la luminosidad instantánea del LHC, y en particular con el cruce de haces en el detector. Se han estudiado las zonas que sufren mayor radiación de fondo, y los resultados se han extrapolado a condiciones de más alta luminosidad, como las alcanzadas durante 2012 y las que se esperan cuando el LHC comience a funcionar otra vez. La conclusión (confirmada en 2012) fue que las tasas de ruido de fondo estarían bajo control y no implicarían ningún problema en la reconstrucción de muones ni en la Física involucrada.
A continuación, el capítulo 8 detalla el análisis $H \to ZZ \to l^+l^-q\bar{q}$. Dos análisis independientes, aunque similares, se realizaron con las muestras de datos recogidas a 7 TeV y 8 TeV. La enorme contribución de fondo de procesos Z+jets, así como otras contribuciones, como $t\bar{t}$ o la producción de dibosones en el SM, obliga a optimizar la selección de sucesos, basada en las diferencias entre los procesos de señal y de fondo, en un intento de mejorar la razón de señal sobre el fondo S/B. Aquí se incluyen diferencias cinemáticas de los objetos involucrados en el decaimiento, así como correlaciones de espín y paridad.

Con el objeto de mejorar la sensibilidad del análisis, la muestra se ha dividido de acuerdo con el número de chorros hadrónicos identificados como procedentes de quarks b. Cada submuestra ha sido tratada separadamente.

Se ha determinado independientemente la contribución de cada proceso de fondo. La más pequeña, que es la producción de pares de bosones, se estima a partir de la simulación, mientras que el $t\bar{t}$ se calcula a partir de una muestra de control a partir de datos $e^{\pm}\mu^{\mp}+X$ con un electrón y un muon. El proceso Z+jets dominante se determina a partir de una región de control en el espectro de m_{jj}, y a partir de la simulación de MC. Toda la estrategia del análisis se ha establecido estudiando las regiones de control y la simulación, para evitar ningún sesgo de los datos de señal en las conclusiones obtenidas.

Los resultados, que incluyen la evaluación de todas las incertidumbres sistemáticas relevantes asociadas al análisis, se presentan en el capítulo 10 a partir de una

interpretación frecuentista. Se han obtenido a partir de un total de 5 fb^{-1} de datos a $\sqrt{s}$ =7 TeV y de 19.6 fb^{-1} a $\sqrt{s}$ =8 TeV. No se ha observado ningún exceso de sucesos por encima de la predicción del fondo en el rango de estudio, y se ha excluido la producción de un bosón de Higgs del SM al 95% de nivel de confianza en el rango de masas entre 275−600 GeV. Aunque no se han establecido límites a la producción de otros modelos, no se ha detectado ningúna indicación que sugiera ninguno de ellos. Finalmente, los pasos a seguir a partir de los resultados presentados incluyen la extensión del análisis para hipótesis de masa de hasta 1 TeV. Ello implica estudiar el caso en que los dos chorros hadrónicos se solapen y aparezcan como un único chorro en el detector. Además, un análisis especialmente dedicado al modo de producción VBF mejorará la sensibilidad del canal. Y, en tercer lugar, extender la interpretación del SM a otros modelos, como Supersimetría o singletes electrodébiles. Todas estas mejoras se esperan en un futuro próximo. En un futuro más lejano, el análisis de los datos recogidos a partir de 2015 a una energía en centro de masas mayor (de al menos $\sqrt{s}$=13 TeV) y en condiciones de mayor luminosidad, permitirá aumentar el rango de masas bajo estudio, la sensibilidad del análisis, y aumentar la cantidad de modelos susceptibles de estudio.

Chapter 14

Agradecimientos-Acknowledgements

Los últimos cuatro años y medio han sido una experiencia inolvidable. Poder aprender tanto de gente tan preparada, y participar de un proyecto tan interesante han sido un absoluto privilegio. Durante este tiempo, además de momentos impagables en lo profesional y lo personal, muchos han sido los dolores de cabeza y el esfuerzo necesario para salir adelante. Afortunadamente he contado con el apoyo de muchas personas que me han ayudado un montón. Y sería extremadamente injusto no aprovechar estas líneas para agradecérselo de todo corazón. Aunque como son tantas, seguro que me olvido de alguien sin querer, pido disculpas anticipadas.

En primer lugar, tengo que expresar mi infinita gratitud a dos de las personas más sabias que he conocido en la vida: Mary-Cruz y Óscar. Desde el primer día hasta hoy que escribo estas líneas, han estado en todo momento formándome y ayudándome en todo. Además son unas personas maravillosas. A pesar de mi corta experiencia, dudo que se puedan tener jefes mejores. Sinceramente, muchas gracias.

El resto del equipo del CIEMAT también ha tenido un comportamiento extraordinario. Empezando por Marcos, que siempre ha mostrado interés por mi evolución y me ha ayudado en todo lo posible. Siguiendo por Manuel, y por todos los investigadores de CMS. A Juan, que es como el oráculo de Delfos salvo para las porras de fútbol. A Ignacio, a Cristina y a Jesús, mis maestros de las DT's. Juan Pablo, por sus solos de violín y sus chistes lamentables. A Isabel, Begoña, Carlo, Silvia, Mara, Conchita... gracias.

Los becarios que me precedieron se merecen un monumento, no solo por ser tan buenos y pacientes maestros, si no como incalificables compañeros de despacho: Carmiña, Marialu, Javierito. Gracias por todo.

Y al resto de becarios, ingenieros y gentes de malvivir con quienes compartí despachos, comidas y profundas charlas filosóficas de sobremesa: Carlos, Aurelio, Cristi, Jorge, Antonio, Adri, Jose, Mariano, Manu, Rafa, Julia, Nacho, el otro Nacho, Miguel, Juanjo, Bruno, ... y a los que llegasteis más tarde y apenas compartimos mantel, pero

espero que sigais manteniendo el nivel.

I must also thank all my coleagues at CERN and all the 2l2q team who helped me, supported me and suffered me through these years, specially Matthias, Annapaola, Miguel and Francesca.

A aquellos profesores que me enseñaron lo bonita que era la Física. A Hans, por poner carteles en la facultad anunciando becas predoctorales de un sitio llamado CIEMAT y convencerme de solicitar una. Y por recomendarme de la de CMS (aunque fuese en segundo lugar. Los tontos siempre tienen suerte).

A mis amigos, que siempre han estado ahí. En especial a Luisete y a Darío, que cierta tarde de Peñas de 2009 delante del Salnés, ante mis dudas, me animaron a hacer las maletas irme a Madrid, a descubrir de qué iba aquello de las partículas. A todos los que me han dado cobijo en Madrid en mis idas y venidas: Luis, Charlotte, Jorge, Iris, Migui, Noemi ...

A todos los que me habéis venido a visitar.

He tenido la suerte de encontrar una segunda Familia en Ginebra. Gracias a Jose, Javi, Tito y Garo por los años que compartimos. No tengo palabras. To all LeClub and all their fans; to all who stayed for dinner and cleaned the next day: to Carlos, to Laurinha, Jabuti, Chelito, Irene, Luis, Bruninho, Byron, Claire, Tintxo, Ili, Helin, Cadi, Roger, Javi, Caro, Nani ... you are so many that I cannot write the full list here. To THAT house. Thank you.

A Natacha, a quem devo muitos lanches e infinitos cafés. Nadge et Yesika, pour votre gentillesse et aide avec des visites guidées.

A Susana, que piensa en todo el mundo menos en ella. Por haberme hecho un regalo tan grande.

A la persona que más me ha apoyado, que más me ha exigido, que más paciencia ha tenido conmigo y en resumen, que más ha hecho para llegar aquí: Gracias Isa.

Finalmente, á toda a miña familia, que seguen enganados pensando que son un fillo modelo. Graciñas por ter tanta confianza en min, totalmente inmerecida. E especialmente ós meus pais, que merecen tódolos títulos de doutor do mundo, que se esforzaron por chegarmos ata aquí moito máis ca min e que son o mellor exemplo que alguén podería seguir. Mamá e papá: grazas por todo.

Remato xa. Se houbese dúas persoas que se aledarían aínda máis cós meus pais por verme rematar esta etapa, son dúas que por desgracia non a chegaron tempo de vivilo. Síntoo na alma. Delas recibín tanto, que non podo menos que adicarlles este traballo, e tódolos éxitos que me poidesen vir no futuro.

Gracias por tener tanto que agradecer.

References

[1] J. B. et al. (Particle Data Group) *Phys. Rev. D* **86, 010001** (2012).

[2] D. Domínguez-Vázquez, *Performance of the Drift Chambers of the CMS Experiment in the Measurement of LHC Muons*, Master's thesis, CIEMAT, 2010.

[3] CMS Collaboration, *The performance of the CMS muon detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV at the LHC*, *JINST* **8** (2013) P11002.

[4] CMS Collaboration, *Search for a Higgs boson in the decay channel H→ ZZ→ qqll in pp collisions at* $\sqrt{s}$*=7 TeV*, *JHEP* **04** (2012) 036.

[5] CMS Collaboration, *Search for a Standard Model-like Higgs boson in the H→ ZZ→ 2l2q channel at* $\sqrt{s}$*= 8 TeV*, .

[6] CMS Collaboration, *Search for a standard-model-like Higgs boson with a mass in the range 145 to 1000 GeV at the LHC*, *Eur. Phys. J. C*.

[7] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* (1967), no. 19.

[8] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[9] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984.

[10] Nobelprize.org, *Press Release: The 1976 Nobel Prize in Physics*, 2013.

[11] UA1 collaboration, *Experimental observation of large transverse energy electrons with associated missing energy at* $\sqrt{s}$*= 540 GeV*, *Physics Letters* **122** (1983) 103–116.

[12] UA2 collaboration, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, .

[13] UA1 collaboration, *Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider*, *Physics Letters* **126** (1983) 398–410.

[14] UA2 Collaboration, *Evidence for $Z^0 \to e^+e$ at the CERN pp collider*, *Physics Letters* **129** (1983) 130–140.

[15] CDF Collaboration, *Observation of Top Quark Production in pp Collisions with the Collider Detector at Fermilab*, *Physical Review Letters* **74** (1995), no. 14 26262631.

[16] D0 Collaboration, *Search for High Mass Top Quark Production in pp Collisions at $\sqrt{s}s$ = 1.8 TeV*, *Physical Review Letters* **74** (1995), no. 13 24222426.

[17] DONUT Collaboration, *Observation of tau neutrino interactions*, *Physics Letters B* **504** (2001), no. 3 218.

[18] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Physics Letters B* (2012), no. 706.

[19] ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, *Physics Letters B* (2012), no. 706.

[20] E. Noether and M. A. Tavel, *Invariant variation problems*, *arXiv:physics/0503066*.

[21] S. L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics*.

[22] A. Salam, *Weak and electromagnetic interactions*, *Elementary particle physics: relativistic groups and analyticity* (1968). Proceedings of the eighth Nobel symposium.

[23] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* (1964), no. 13.

[24] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* (1964), no. 12.

[25] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* (1964), no. 13.

[26] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, *Phys. Rev.* (1966), no. 145.

[27] R. Hagen and T. W. B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* (1964), no. 13.

[28] T. W. B. Kibble, *Symmetry breaking in non-Abelian gauge theories*, *Phys. Rev.* (1967), no. 155.

[29] ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD Collaborations and the LEP Electroweak Working Group and the Tevatron Electroweak Working Group and the SLD Electroweak and Heavy Flavour Groups, *Precision Electroweak Measurements and Constraints on the Standard Model*, .

[30] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, *CERN-2011-002* (CERN, Geneva, 2011) [arXiv:1101.0593].

[31] S. Goria, G. Passarino, and D. Rosco, *The Higgs Boson Lineshape*, *Nucl.Phys.* **B864** (2012) 530–579, [arXiv:1112.5517].

[32] CMS Collaboration, *Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel*, *CMS Physics Analysis Summary CMS-PAS-HIG-13-001* (2013).

[33] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, *Phys. Lett. B* **726** (2013), no. CERN-PH-EP-2013-103 88.

[34] CMS Collaboration, *Properties of the Higgs-like boson in the decay H to ZZ to 4l in pp collisions at sqrts =7 and 8 TeV*, *CMS Physics Analysis Summary CMS-PAS-HIG-13-002* (2013).

[35] CMS Collaboration, *Evidence for a particle decaying to W^+W^- in the fully leptonic final state in a standard model Higgs boson search in pp collisions at the LHC*, *CMS Physics Analysis Summary CMS-PAS-HIG-13-003* (2013).

[36] CMS Collaboration, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, *Phys. Rev. D* (2013), no. CMS-HIG-13-012, CERN-PH-EP-2013-188.

[37] CMS Collaboration, *Evidence for a Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV*, .

[38] *Search for the bb decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector*, .

[39] ATLAS Collaboration, *Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector*, .

[40] CMS Collaboration, *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*, *CMS Physics Analysis Summary CMS-PAS-HIG-13-005* (2013).

[41] ATLAS Collaboration, *Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^1 of proton-proton collision data*, *ATLAS-CONF-2013-014*.

[42] CMS Collaboration, *Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs*, *Physics Review Letters* (2013), no. 110.

[43] ATLAS Collaboration, *Study of the spin of the new boson with up to 25 fb^1 of ATLAS data*, *ATLAS-CONF-2013-034*.

[44] CMS Collaboration, *Properties of the observed Higgs-like resonance using the diphoton channel*, *CMS Physics Analysis Summary CMS-PAS-HIG-13-016* (2013).

[45] P. Binetruy, *Supersymmetry: Theory, Experiment, and Cosmology*. Oxford University Press, 2006.

[46] S. P. Martin, *A Supersymmetry Primer*, *arXiv:hep-ph/9709356v5* (2008).

[47] S. LHC Higgs Cross Section Working Group, Heinemeyer *et al.*, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, `arXiv:1307.1347`.

[48] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation*.

[49] O. Bruning *et al.*, *The LHC design report v.1: the LHC Main Ring*, *CERN-2004-003-V-1* (2004).

[50] O. Bruning *et al.*, *The LHC design report v.2 : the LHC Infrastructure and General Services*, *CERN-2004-003-V-2* (2004).

[51] B. Benedikt *et al.*, *The LHC design report v.3 : the LHC Infrastructure and General Services*, *CERN-2004-003-V-3* (2004).

[52] CERN, *CONVENTION FOR THE ESTABLISHMENT OF A EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH*, 1953.

[53] LHCb Collaboration, *LHCb technical design report: Reoptimized detector design and performance*, *CERN-LHCC-2003-030* (2003).

[54] ALICE Collaboration, *ALICE: Technical proposal for a large ion collider experiment at the CERN LHC*, *CERN-LHCC-95-71* (1995).

[55] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST 3 S08003* (2008).

[56] CMS Collaboration, *CMS technical proposal*, *CERN-LHCC-94-38* **http://cdsweb.cern.ch/record/290969.**

[57] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST 0803:S08004* (2008).

[58] TOTEM Collaboration, *Total cross-section, elastic scattering and diffraction dissociation at the large hadron collider at cern : Totem technical design report*, *CERN-LHCC-2004-002* (2004).

[59] TOTEM Collaboration, *Total cross-section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN : Addendum to the totem technical design report*, *CERN-LHCC-2004-020* (2004).

[60] LHCf Collaboration, *LHCf experiment : Technical Design Report*, *CERN-LHCC-2006-004* (2006).

[61] The MoEDAL Collaboration, *Technical design report of the MoEDAL experiment*, *CERN-LHC-2009-006* (2010).

[62] CMS Collaboration, *CMS Physics Technical Design Report*, *CERN/LHCC 2006-001 2178*.

[63] CMS Collaboration, *The CMS tracker system project: Technical Design Report*, *CERN/LHCC* **98-006** (1998). CMS TDR 5.

[64] CMS Collaboration, *The CMS Electromagnetic Calorimeter Project : Technical Design Report*, .

[65] CMS Collaboration, Chatrchyan, Serguei and others, *Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam, Cosmic Ray, and LHC Beam Data*, *JINST* **5** (2010) T03013, [arXiv:0911.4877].

[66] CMS Collaboration, *CMS MUON Technical Design Report*, *CERN/LHCC* **32** (1997).

[67] CMS Collaboration, *The performance of the CMS muon detector in protonproton collisions at* $\sqrt{s}$ *= 7 TeV at the LHC*, *JINST* **8** (2013) P11002.

[68] CMS Collaboration, *CMS trigger and data-acquisition project : Technical Design Report*, .

[69] CMS Collaboration, *The Level-1 Trigger Technical Design Report*, *CERN/LHCC 2164 2000-038* (2000).

[70] R. Brun and F. Rademakers, *ROOT - an object oriented data analysis framework*, *Inst. Meth. in Phys. Res. A* **389** (1997) 81.

[71] GEANT4 Collaboration, Agostinelli, S. and others, *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.

[72] J. P. Pelayo, *Estudio sobre las Cámaras de Tubos de Deriva para el Espectrometro de Muones del Experimento CMS*. PhD thesis, Universidad Autónoma de Madrid, 2004.

[73] CMS Collaboration, *Calibration of the CMS Drift Tube Chambers and Measurement of the Drift Velocity with Cosmic Rays*, *JINST* **5** (2010) T03016.

[74] *https://twiki.cern.ch/twiki/bin/view/CMSPublic/MuonDPGResults*. Muon DPG Approved Results.

[75] CMS Collaboration, *Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and Emiss*, *CMS-PAS-PFT-09-001* (2009).

[76] CMS Collaboration, *Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV*, *CMS-PAS-PFT-10-002* (2010).

[77] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual (for version 3.0.2)*, *Eur. Phys. J. C* **72** (2012) 1896, [arXiv:1111.6097].

[78] CMS Collaboration, *Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$=7 TeV*, *JINST* **8** (2013) P09009.

[79] CMS Collaboration, *Electron performance with 19.6 fb^1 of data collected at $\sqrt{s}$=8 TeV with the CMS detector*, *CMS-DP-2013-003* (2013).

[80] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189].

[81] CMS Collaboration, Chatrchyan, Serguei and others, *Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS*, *JINST* **6** (2011) P11002, [arXiv:1107.4277].

[82] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, *J. Instrum* (2013).

[83] CMS Collaboration, *Search for a heavy Higgs boson in the H$\rightarrow$ ZZ $\rightarrow$ 2l2ν channel in pp collisions at $\sqrt{s}$= 7 and 8 TeV*, CMS Physics Analysis Summary CMS-PAS-HIG-13-014, 2013.

[84] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : going beyond*, *JHEP* **06** (2011) 128, [arXiv:1106.0522].

[85] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, *JHEP* **11** (2004) 040.

[86] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD Computations with Parton Shower Simulations: the POWHEG method*, *JHEP* **11** (2007) 070.

[87] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO Vector-Boson Production Matched with Shower in POWHEG*, *JHEP* **07** (2008) 06.

[88] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[89] The CMS generator group, *https://twiki.cern.ch/twiki/bin/viewauth/cms/standardmodelcrosssectionsat8tev*, twiki, CERN, 2012.

[90] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, *et al.*, *New parton distributions for collider physics*, *Phys.Rev.* **D82** (2010) 074024, [arXiv:1007.2241].

[91] S. Kretzer, H. Lai, F. Olness, and W. Tung, *CTEQ6 parton distributions with heavy quark mass effects*, *Phys. Rev. D* **69** (2004) 114005, [hep-ph/0307022].

[92] LHC Higgs Cross Section Working Group web page, *https://twiki.cern.ch/twiki/bin/view/LHCPhysics*.

[93] Y. Gao *et al.*, *Spin determination of single-produced resonances at hadron colliders*, *Phys. Rev.* **D81** (2010) 075022.

[94] N. Cabibbo and A. Maksymowicz, *Angular Correlations in Ke-4 Decays and Determination of Low-Energy pi-pi Phase Shifts*, *Phys.Rev.* **137** (1965) B438–B443.

[95] N. Kauer and G. Passarino, *Inadequacy of zero-width approximation for a light Higgs boson signal*, *JHEP* **08** (2012) 116, [arXiv:1206.4803].

[96] G. Passarino, *Higgs Interference Effects in $gg \to ZZ$ and their Uncertainty*, *JHEP* **08** (2012) 146, [arXiv:1206.3824].

[97] CMS Collaboration, *CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update*, CMS Physics Analysis Summary CMS-PAS-LUM-12-001, CERN, 2012.

[98] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, *Eur.Phys.J.* **C63** (2009) 189–285, [arXiv:0901.0002].

[99] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, *et al.*, *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, *Nucl.Phys.* **B849** (2011) 296–363, [arXiv:1101.1300].

[100] N. Kauer, *Signal-background interference in $gg \to H \to VV$*, arXiv:1201.1667.

[101] CMS Collaboration, ATLAS Collaboration, and LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in summer 2011*, *ATL-PHYS-PUB-2011-011, CMS NOTE-2011/005* (2011).

[102] L. Moneta *et al.*, *The RooStats Project*, *ACAT2010 Conference Proceedings* (2010) [arXiv:1009.1003].

[103] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl.Instrum.Meth.* **A434** (1999) 435–443, [hep-ex/9902006].

[104] A. L. Read, *Presentation of search results: the CLs technique*, *J. Phys. G: Nucl. Part. Phys.* **28** (2002).

[105] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur.Phys.J.* **C71** (2011) 1554, [arXiv:1007.1727].